ELSEVIER

Chemometrics and Intelligent Laboratory Systems 55 (2001) 67–90

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Iteratively reweighted generalized rank annihilation method 1. Improved handling of prediction bias

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

[a] *Dunantsingel 28, NL-2806 JB Gouda, Netherlands*

[b] *Department of Analytical and Organic Chemistry, Institute of Advanced Studies, Rovira 1 Virgili University, Pça. Imperial Tàrraco 1, E-43005 Tarragona, Catalonia, Spain*

Received 13 April 2000; accepted 13 November 2000

Abstract

The generalized rank annihilation method (GRAM) is a method for curve resolution and calibration that uses two bilinear matrices simultaneously, i.e., one for the unknown and one for the calibration sample. A GRAM calculation amounts to solving an eigenvalue problem for which the eigenvalues are related to the predicted analyte concentrations. Previous studies have shown that random measurement errors bring about a bias in the eigenvalues, which directly translates into prediction bias. In this paper, accurate formulas are derived that enable removing most of this bias. Two bias correction methods are investigated. While the first method directly subtracts bias from the eigenvalues obtained by the original GRAM, the second method first applies a weight to the data matrices to reduce bias. These weights are specific for the analyte of interest and must be determined iteratively from the data. Consequently, the proposed modification is called iteratively reweighted GRAM (IRGRAM). The results of Monte Carlo simulations show that both methods are effective in the sense that the standard error in the bias-corrected prediction compares favourably with the root mean squared error (RMSE) that accompanies the original quantity. However, IRGRAM is found to perform best because the increase of variance caused by subtracting bias is minimised. In the original formulation of GRAM only a single calibration sample is exploited. The error analysis is extended to cope with multiple calibration samples.

Keywords: Bilinear calibration; GRAM; IRGRAM; Prediction bias; Bias correction

1. Introduction

The generalized rank annihilation method (GRAM) enables the prediction of analyte concentration in the presence of unknown interferences [1]. From the analytical perspective, it is desirable that the predicted analyte concentrations be labelled with a measure of their uncertainty. However, to assess this uncertainty, one has to investigate how measurement errors propagate through the GRAM eigenvalue problem, which is cumbersome owing to the complexity of the algebra involved. While most researchers have focused on prediction variance (see Ref. [2] and references therein), Booksh and Kowalski [3] and Faber et al. [4] have shown that random measurement errors also bring about a prediction bias.

* Corresponding author. ATO, Department Production and Control Systems, P.O. Box 17, 6700 AA Wageningen, Netherlands.

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

0169-7439/01/$ - see front matter
PII: S0169-7439(00)00117-9

(Variance quantifies the spread of an individual prediction around the mean or expected value; in contrast, bias constitutes a systematic deviation of the mean value from the true value.) Faber et al. [4] have derived expressions for estimating and correcting bias in the eigenvalues which seem to work well if the bias does not exceed 15%. However, the Monte Carlo simulations performed by Booksh and Kowalski [3] clearly demonstrate that bias may be as large as 20%. Consequently, bias correction as proposed by Faber et al. [4] is not generally applicable.

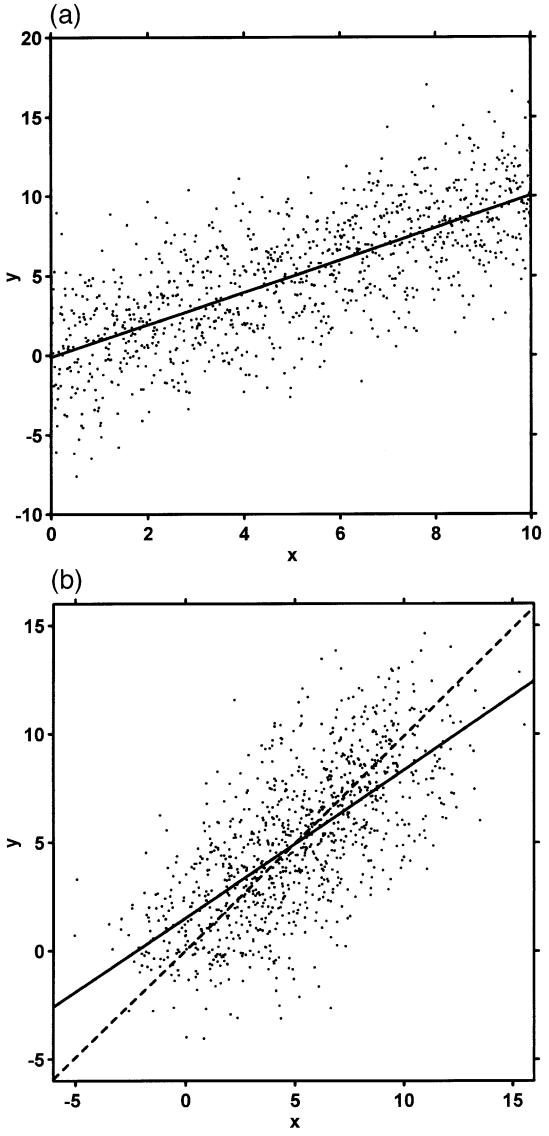

Fig. 1. Straight-line fitting with (a) zero and (b) substantial error in the *x*-variable. The true intercept and slope are zero and unity, respectively. In the first case, the line estimated using OLS (—) is unbiased whereas in the second case a bias-correction is required. For the current example, a successful bias-correction is obtained using corrected least squares (- - -).

It is emphasised that the bias encountered in GRAM is a general phenomenon. It can, for example, be compared to the bias in the regression coefficients estimated by ordinary least squares (OLS) when the predictors carry a non-negligible error. A correction for bias is required to improve the regression coefficient estimates. This is illustrated in Fig. 1 where the results of fitting a straight line using OLS and corrected least squares [5] are compared. For a recent example in analytical chemistry, see Riu and Rius [6] who have studied this problem in the context of method comparison.

The objective of this study is to develop an effective bias correction procedure. The adequate handling of bias in the eigenvalues of GRAM would constitute an important step towards the construction of prediction interval and limit of detection (LOD) estimators. Currently, these estimators are non-existent for the analytical problem at hand, namely, the determination of analyte concentration in the presence of unknown interferences.

2. Notation and terminology

The following notation and terminology are maintained throughout Parts 1 and 2 of this series. To simplify the presentation, additional notation for showing a dependency on the weight factor is avoided throughout. The identity of the analyte of interest is not made explicit either.

2.1. General

Boldface uppercase letters represent matrices, e.g., **A**. The $n \times n$ identity matrix and $n \times m$ null matrix are denoted by $\mathbf{I}_n$ and $\mathbf{O}_{nm}$, respectively. Column vectors will be indicated by boldface lowercase letters, e.g., **a**. Scalars are indicated by italic uppercase or lowercase letters, e.g., *A* and *a*. Transposition of a matrix or vector is symbolised by a superscripted

'T', e.g., $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{a}^{\mathrm{T}}$. The Euclidean norm of a vector or matrix is symbolised by $\|\bullet\|$. The trace and the (principal) diagonal of a matrix are indicated by $\mathrm{Tr}(\cdot)$ and $\mathrm{diag}(\cdot)$, respectively. Vectorisation of a matrix (i.e., stacking its columns from left to right) is indicated by 'vec'. The Kronecker product is symbolised by '$\otimes$'. For a given matrix $\mathbf{A}$, the matrices $\mathbf{A}^{-1}$ and $\mathbf{A}^{+}$ stand for its inverse and (unique) Moore–Penrose pseudoinverse, respectively. The 'inverse transpose' and 'pseudoinverse transpose' of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^{-\mathrm{T}}(=(\mathbf{A}^{-1})^{\mathbf{T}}=(\mathbf{A}^{\mathrm{T}})^{-1})$ and $\mathbf{A}^{+\mathrm{T}}(=(\mathbf{A}^{+})^{\mathrm{T}}=(\mathbf{A}^{\mathrm{T}})^{+})$, respectively. Measured quantities are distinguished from their errorless counterparts (true values) by adding a 'tilde', e.g., $\tilde{\mathbf{A}}$. It is important to note that the notation for true and measured quantities has changed relative to Ref. [4]. Likewise, estimated or predicted quantities carry a 'hat', e.g., $\hat{\mathbf{A}}$. Expected value, variance, standard deviation, bias and root mean squared error (RMSE) are denoted by $E(\cdot)$, $V(\cdot)$, $\sigma(\cdot)$, $B(\cdot)$ and $RMSE(\cdot)$, respectively.

2.2. Net analyte signal

It has been shown that the variance expressions originally derived for GRAM [2,4,7] can be simplified considerably using consistently defined analytical figures of merit for bilinear matrix data [8]. These figures of merit are based on a suitable generalisation of Lorber's scalar net analyte signal (NAS) for vector data [9] (i.e., multivariate calibration). The superscripted '*' is used to symbolise the NAS, e.g., r^{*}.

2.3. The method of error propagation for obtaining approximate expressions for bias and variance

The method of error propagation deals with the way uncertainties in the input data are carried over or propagated to the uncertainty in the final result of a calculation (estimate or prediction). The rationale can be understood by considering the bias and variance in the linear function $a\tilde{X}+b\tilde{Y}$ where a and b are constants and $\tilde{X}$ and $\tilde{Y}$ random variables with expectations X and Y and variances $V(\tilde{X})$ and $V(\tilde{Y})$, respectively. The exact bias is zero because $B(a\tilde{X}+b\tilde{Y})\equiv E(a\tilde{X}+b\tilde{Y})-aX-bY=aE(\tilde{X})+bE(\tilde{Y})-aX-bY=0$. Moreover, the *exact* variance is given by $V(a\tilde{X}+b\tilde{Y})=a^2V(\tilde{X})+b^2V(\tilde{Y})$. For simple non-linear functions, such as a χ^2 random variable, it is often possible to obtain exact bias and variance expressions by calculating moments of its distribution. Unfortunately, the GRAM eigenvalue $\hat{\pi}$ is a complicated non-linear function, hence, its distribution is untractable and exact results for bias and variance are out of reach.

The method of error propagation handles this non-linearity by expanding the eigenvalue as $\hat{\pi}=\pi+\mathrm{d}\pi+\mathrm{d}^2\pi+\ldots$ where the prefix symbol d^n is used to symbolise the nth differential. The first differential constitutes a linear approximation of $\hat{\pi}$ while the higher-order terms describe the curvature. Thus, by considering one term beyond the linear one, an approximate bias for $\hat{\pi}$ follows as

$$\begin{aligned} B(\hat{\pi}) &\equiv E(\hat{\pi})-\pi \\ &\approx E(\pi+\mathrm{d}\pi+\mathrm{d}^2\pi)-\pi \\ &= E(\mathrm{d}^2\pi) \end{aligned} \tag{1}$$

where it has been used that $E(\mathrm{d}\pi)=0$, which holds because $\mathrm{d}\pi$ is a linear function of the input data (see Eq. (21) in Part 2 [10]), which are assumed to be unbiased.

Likewise, an approximate variance for $\hat{\pi}$ is obtained by truncating after the first-order term, which is often referred to as *local linearisation*, i.e.,

$$\begin{aligned} V(\hat{\pi}) &\equiv E\left((\hat{\pi}-E(\hat{\pi}))^2\right) \\ &\approx E\left((\pi+\mathrm{d}\pi-E(\pi+\mathrm{d}\pi))^2\right) \\ &= E\left((\mathrm{d}\pi)^2\right) \end{aligned} \tag{2}$$

In summary, the method of error propagation yields approximations to bias and variance by truncating a series expansion. While the variance expression follows from a first-order approximation, a second-order approximation is required to obtain a bias expression. Vectors and matrices are handled in straightforward fashion by applying similar expressions to the individual elements [11]. The adequacy of the obtained bias or variance expression is conveniently tested using Monte Carlo simulations.

3. Theory

In this section, only the main results of the derivations are presented. The expression for determining the bias in the eigenvalues is derived in Appendix A while Appendix B deals with the uncertainty in the formula-based bias estimates. For convenience, it is assumed throughout that the measurement errors are independently and identically distributed (iid). The handling of heteroscedastic and correlated errors, which is the practical condition, is outlined in Appendix C. The fact that the generalization of the derived expressions is straightforward can be seen as an indication that the currently developed error analysis is more than an exercise in futility. A simulation technique, based on adding noise to the data, is detailed in Appendix D. This technique is useful for assessing the adequacy of approximate expressions in practice.

3.1. Preliminaries

Accurate analyte determination requires that the predicted value, $\hat{c}_u$, falls close to the true value, c_u. A quantitative measure for the expected deviation from the true value is given by the RMSE, which has a variance and bias contribution:

$$\begin{aligned} RMSE(\hat{c}_u) &\equiv \sqrt{E\left((\hat{c}_u - c_u)^2\right)} \\ &= \sqrt{E\left((\hat{c}_u - E(\hat{c}_u))^2\right) + (E(\hat{c}_u) - c_u)^2} \\ &= \sqrt{V(\hat{c}_u) + B(\hat{c}_u)^2} \end{aligned} \quad (3)$$

Eq. (3) suggests a natural criterion for assessing the effectiveness of bias correction. Correction for bias leads to an increased standard error (square root of the variance) because a number is subtracted that carries an uncertainty. Thus, a bias correction method can only be called effective if the standard error in the bias-corrected quantity compares favourably with the RMSE that characterises the original quantity (see Fig. 2). For the reason of assessing the effectiveness of bias corrections, considerable attention will be given to variance and standard error in the remainder of this paper.

3.2. Gram eigenvalue problem

The input of a GRAM calculation consists of $J_1 \times J_2$ bilinear data matrices,

$$\begin{aligned} \tilde{\mathbf{R}}_0 &= \mathbf{X}\mathbf{C}_0\mathbf{Y}^T + \mathbf{E}_0 \\ \tilde{\mathbf{R}}_u &= \mathbf{X}\mathbf{C}_u\mathbf{Y}^T + \mathbf{E}_u \end{aligned} \quad (4)$$

where $\tilde{\mathbf{R}}_0$ and $\tilde{\mathbf{R}}_u$ are measured for the calibration and unknown sample, $\mathbf{X}$ ($J_1 \times K$) and $\mathbf{Y}$ ($J_2 \times K$) contain the column and row profiles (K is the number of constituents), $\mathbf{C}_0$ and $\mathbf{C}_u$ are $K \times K$ diagonal matrices of concentrations, and $\mathbf{E}_0$ and $\mathbf{E}_u$ are $J_1 \times J_2$ error matrices. For convenience, it will be assumed throughout that the elements of $\mathbf{E}_0$ and $\mathbf{E}_u$ are independently and identically distributed (iid) with constant variance $V(\tilde{\mathbf{R}}_0) = V(\tilde{\mathbf{R}}_u) = V(\tilde{\mathbf{R}})$. In addition, it is assumed that $\mathbf{X}$ and $\mathbf{Y}$ are full column rank matrices. Kiers and Smilde [12] give a separate treatment of the case where some of the columns of $\mathbf{X}$ and $\mathbf{Y}$ are completely overlapped with the remaining ones. The current analysis still holds for the analyte of interest if it is not involved in a so-called rank overlap. For the results to carry through, one must replace the profiles of the constituents that are involved in a rank overlap by a correspondingly smaller number, and adapt $\mathbf{C}_0$ and $\mathbf{C}_u$ accordingly. This can be done without changing the notation. (For example, K now denotes the number of constituents minus the number of rank overlaps.)

GRAM is a principal component analysis (PCA)-based curve resolution method: In the first step, orthogonal factors (scores $\mathbf{S}$ and loadings $\mathbf{L}$) are estimated by decomposing a data matrix using the singular value decomposition (SVD). In the second step, which amounts to solving an eigenvalue problem, these factors are transformed to obtain estimates for the column and row profiles ($\mathbf{X}$ and $\mathbf{Y}$). The main difference with other PCA-based curve resolution methods (e.g., evolving factor analysis or iterative target testing factor analysis) is that the transformation step involves a second data matrix, namely, from the calibration sample. This difference enables one to obtain the quantitative information as well, since the eigenvalues are directly related to analyte concentrations in the unknown sample.

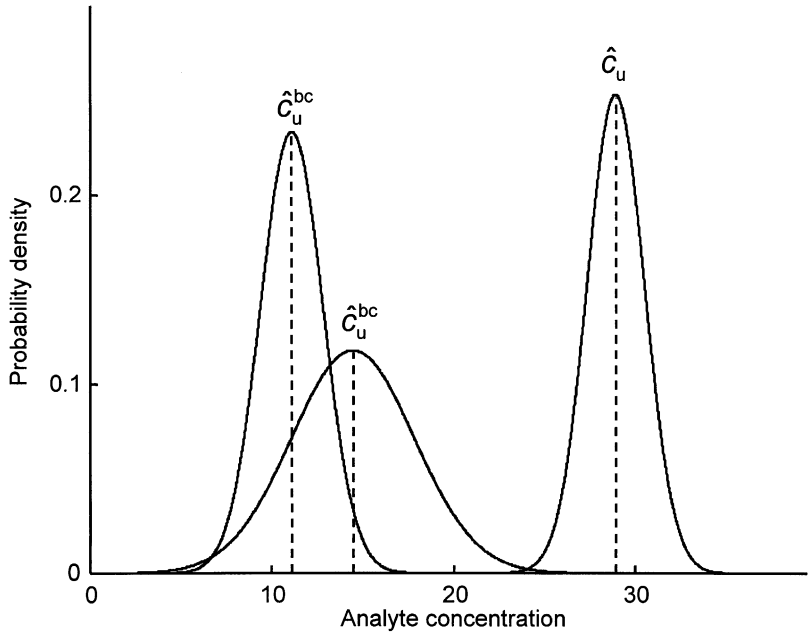

Fig. 2. Illustration that bias-correction yields an increase of standard error. (This numerical example will be treated in detail in the Results and discussion section.) The probability density function of the original prediction ($\hat{c}_u = 28.96 \pm 1.58$) is narrower than those for the two bias-corrected predictions ($\hat{c}_u^{bc} = 11.10 \pm 1.71$ and $\hat{c}_u^{bc} = 14.46 \pm 3.39$). Since the true analyte concentration (c_u) is 10, both bias corrections are successful in the sense that the increased variance is more than offset by the reduction of bias. However, one bias correction can be regarded as more effective because the increase of the standard error is smaller.

In the general case, both samples contain substituents that are not present in the other one. Consequently, neither $\tilde{\mathbf{R}}_0$ nor $\tilde{\mathbf{R}}_u$ alone can be used for the construction of the factor space. Sánchez and Kowalski [1] have suggested decomposing the sum matrix $\tilde{\mathbf{Q}} = \tilde{\mathbf{R}}_u + \tilde{\mathbf{R}}_0$, and projecting $\tilde{\mathbf{R}}_u$ onto the factor space to obtain the desired resolution for the analyte of interest. Our proposed modification simply amounts to decomposing $\tilde{\mathbf{Q}} = \tilde{\mathbf{R}}_u + \alpha\tilde{\mathbf{R}}_0$, where $\alpha \neq 0$[1]. It is noted that Leurgans et al. [13] discuss a bilinear calibration method similar to GRAM where weights can be applied to $\tilde{\mathbf{R}}_0$ and $\tilde{\mathbf{R}}_u$. They also state a requirement on algebraic grounds, but choose the value unity for the actual calculations. (Here the weight for $\tilde{\mathbf{R}}_u$ is unity.) Only the main steps leading to the modified GRAM eigenvalue problem are given here to introduce the necessary quantities. For more details about GRAM, see Refs. [1,12].

For what follows, it is convenient to express the relevant data matrices $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}_u$ as

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{R}}_u + \alpha\tilde{\mathbf{R}}_0 = \mathbf{H}\mathbf{Y}^T + \mathbf{E}_Q$$

$$\tilde{\mathbf{R}}_u = \mathbf{H}\mathbf{\Pi}\mathbf{Y}^T + \mathbf{E}_u \quad (5)$$

where $\mathbf{H} = \mathbf{X}(\mathbf{C}_u + \alpha\mathbf{C}_0)$ and $\mathbf{\Pi} = (\mathbf{C}_u + \alpha\mathbf{C}_0)^{-1}\mathbf{C}_u$.

The SVD of $\tilde{\mathbf{Q}}$ yields

$$\tilde{\mathbf{Q}} = \hat{\mathbf{U}}\hat{\mathbf{\Theta}}\hat{\mathbf{V}}^T + \hat{\mathbf{E}}_Q \quad (6)$$

where $\hat{\mathbf{U}}$ $(J_1 \times F)$ and $\hat{\mathbf{V}}$ $(J_2 \times F)$ contain the left and right singular vectors, $\hat{\mathbf{\Theta}}$ is the $F \times F$ diagonal matrix of singular values in non-increasing order, and $\hat{\mathbf{E}}_Q$ denotes the residual matrix from the SVD model fit of $\tilde{\mathbf{Q}}$ (only F significant factors are retained; F is an estimate of K). The SVD provides a convenient way to perform a PCA, since the score and loading matrix

[1] Since $\tilde{\mathbf{R}}_u$ takes part in the decomposition as well as the projection step, its weight is arbitrary as long as the weight for $\tilde{\mathbf{R}}_0$ is adjusted accordingly. To simplify the presentation, $\tilde{\mathbf{R}}_u$ has received the weight unity.

can be expressed as $\mathbf{S} = \mathbf{U}\mathbf{\Theta}$ and $\mathbf{L} = \mathbf{V}^{\mathrm{T}}$, respectively. Faber et al. [14] have listed several variations of the GRAM eigenvalue problem. The following form,

$$\left(\hat{\mathbf{\Theta}}^{-1}\hat{\mathbf{U}}^{\mathrm{T}}\tilde{\mathbf{R}}_{\mathrm{u}}\hat{\mathbf{V}}\right)\hat{\mathbf{T}} = \hat{\mathbf{T}}\hat{\mathbf{\Pi}} \tag{7}$$

is particularly attractive because it gives simple transformation formulas for the factors resulting from SVD, i.e.,

$$\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\mathbf{\Theta}}\hat{\mathbf{T}}$$

$$\hat{\mathbf{Y}} = \hat{\mathbf{V}}\hat{\mathbf{T}}^{-\mathrm{T}} \tag{8}$$

It is seen that the $F \times F$ eigenvector matrix $\hat{\mathbf{T}}$ transforms the score matrix $\hat{\mathbf{S}} = \hat{\mathbf{U}}\hat{\mathbf{\Theta}}$, while its 'inverse transpose' transforms the transpose loading matrix $\hat{\mathbf{V}}$.

The predicted analyte concentration, $\hat{c}_{\mathrm{u}}$, is obtained as

$$\hat{c}_{\mathrm{u}} = \hat{\alpha}\tilde{c}_0\hat{\pi}/(1-\hat{\pi}) \tag{9}$$

where $\tilde{c}_0$ and $\hat{\pi}$ denote the measured analyte concentration in the calibration sample and the associated eigenvalue, respectively. Estimation of α is a key topic in the current paper.

Application of Eq. (9) implies that individual eigenvalues can be identified as belonging to a particular analyte. In practice, this is achieved by comparing the estimated profiles (e.g., the spectrum) with a reference profile [1].

3.3. Bias in the eigenvalues and predictions

The purpose of deriving bias expressions is to enable improving the predictions by correcting for bias. It is shown in Appendix A that the bias in the eigenvalue $\hat{\pi}$ is given by

$$B(\hat{\pi}) \approx \omega\psi\left[1-(1+\alpha^2)\pi\right]V(\tilde{\mathbf{R}}) \tag{10}$$

where $\omega = J_1^{-1}(J_1-K-1)(J_1-K) + J_2^{-1}(J_2-K-1)(J_2-K)$ and ψ is the associated diagonal element of $\mathbf{\Psi} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}$. This expression is more accurate than the one given in Ref. [4].

Eq. (10) shows how the bias depends on the characteristics of the data and the weight α, namely:

- size of the matrices and number of constituents: explicitly through ω and implicitly through ψ (ψ tends to decrease with increasing J_1 and J_2 while it tends to increase with increasing K),
- overlap among the profiles: explicitly through ψ,
- amounts of analyte in unknown and calibration sample: explicitly through ψ, since $\mathbf{H} = \mathbf{X}(\mathbf{C}_{\mathrm{u}} + \alpha\mathbf{C}_0)$ and explicitly through the eigenvalue $\pi = c_{\mathrm{u}}/(c_{\mathrm{u}} + \alpha c_0)$,
- level of the noise: bias is proportional to the noise variance,
- the weight α: explicitly through the term $(1+\alpha^2)$, $\mathbf{H}$, and the eigenvalue π.

It will be detailed below how the weight can be utilised to optimise the prediction.

It is noted that the presence of $\mathbf{H}$ and $\mathbf{Y}$ is somewhat misleading in the sense that it suggests that profile estimates for *all* constituents are required for the evaluation of Eq. (10). However, it is shown in Ref. [15] that a diagonal element of $\mathbf{\Psi}$ can be evaluated using information pertaining to the analyte of interest only. In short, the overlaps between the profiles of the analyte of interest ($\mathbf{h}$ and $\mathbf{y}$) and the spaces spanned by the profiles of the interferences (the remaining columns of $\mathbf{H}$ and $\mathbf{Y}$) are the relevant quantities. Clearly, these overlaps do not depend on the individual contribution of the interferences to these spaces.

Starting from $c_{\mathrm{u}} = \alpha c_0\pi/(1-\pi)$, an exact expression for the bias in the predictions is obtained by adding a bias to the true eigenvalues, i.e.,

$$c_{\mathrm{u}} + B(\hat{c}_{\mathrm{u}}) = \alpha c_0 \frac{\pi + B(\hat{\pi})}{1-(\pi + B(\hat{\pi}))} \tag{11}$$

from which it follows that

$$B(\hat{c}_{\mathrm{u}}) = \frac{\alpha c_0 B(\hat{\pi})}{(1-\pi)(1-\pi-B(\hat{\pi}))} \tag{12}$$

Two remarks seem to be in order. First, Eq. (12) holds independent of the method that is used to obtain $\hat{\pi}$. Second, Eq. (12) can be rewritten as

$$B(\hat{c}_{\mathrm{u}}) = \frac{c_{\mathrm{u}} B(\hat{\pi})}{\pi(1-\pi-B(\hat{\pi}))} \tag{13}$$

and it is found that the percentage bias in $\hat{c}_u$ will be greater than the percentage bias in $\hat{\pi}$ if $B(\hat{\pi}) > -\pi$. Unfortunately, Eq. (13) does not give a solution for $c_u = 0$, and is therefore not suitable for actual calculations.

3.4. Variance in the eigenvalues and predictions

It is derived in Part 2 [10] that, assuming that the measurement errors in the calibration sample matrix, $\tilde{\mathbf{R}}_0$, and the unknown sample matrix, $\tilde{\mathbf{R}}_u$, are iid with variance $V(\tilde{\mathbf{R}}_0) = V(\tilde{\mathbf{R}}_u) = V(\tilde{\mathbf{R}})$, the variance in the eigenvalue $\hat{\pi}$ is given by

$$V(\hat{\pi}) \approx (r_Q^*)^{-2}\left[(1-\pi)^2 + \alpha^2\pi^2\right]V(\tilde{\mathbf{R}}) \qquad (14)$$

where r_Q^* is the associated net analyte signal (NAS) in the decomposed matrix $\mathbf{Q}$. The factor $(r_Q^*)^{-2}$ has the interpretation of a variance inflation factor (VIF) for the eigenvalues, because it quantifies the amount of instrumental error propagation. Using Eq. (14), it is derived in Part 2 [10] that the variance in the prediction $\hat{c}_u$ is given by

$$V(\hat{c}_u) \approx s^{-2}(1+h_u)V(\tilde{\mathbf{R}}) + h_u \cdot V(\tilde{c}_0) \qquad (15)$$

where s denotes the sensitivity [8,9] and $h_u = (c_u/c_0)^2$ is the unknown sample leverage [8]. When focussing attention to the instrumental errors, the factor s^{-2} has the interpretation of a VIF for the predictions. Unlike the prediction bias Eq. (10), the prediction variance Eq. (15) does not depend on the weight α (to the order of the approximation, of course).

3.5. Practical evaluation of bias and variance expressions

The practical evaluation of Eqs. (10), (12), (14) and (15) amounts to inserting measured, estimated and predicted quantities. However, the *correct* practical evaluation is essentially different for bias and variance expressions. This can be understood as follows. Bias quantifies the deviation of the mean from the true value, whereas variance is defined as the spread around the mean. As a result, bias expressions contain true values, whereas variance expressions contain expected values. Practically, this amounts to correct evaluation of bias expressions by inserting bias-corrected quantities, whereas variance expressions should be evaluated with quantities that carry a bias.

Thus, the bias in $\hat{\pi}$ should be estimated as

$$\hat{B}(\hat{\pi}) = \hat{\omega}\hat{\psi}\left[1 - (1+\hat{\alpha}^2)\left(\hat{\pi} - \hat{B}(\hat{\pi})\right)\right]\hat{V}(\tilde{\mathbf{R}}) \qquad (16)$$

where $\hat{\omega} = J_1^{-1}(J_1 - F - 1)(J_1 - F) + J_2^{-1}(J_2 - F - 1)(J_2 - F)$ in which $F = \hat{K}$ is the estimated dimensionality of the GRAM model. It is recommended to estimate K by building GRAM models of increasing dimensionality and comparing the results. When increasing F beyond K, the general features of the estimated profiles will remain the same, but their variance will increase. In addition, the predictions will stabilise. Estimating the dimensionality of a GRAM model is much easier than estimating the pseudorank of a *single* data matrix, i.e., its mathematical rank in the absence of noise, because the results of the GRAM calculation provide useful diagnostics.

It is emphasised that only the eigenvalue $\hat{\pi}$ is corrected for bias. The bias factor $\hat{\psi}$ need not be corrected for the bias in $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$, owing to their interdependency (see discussion of term (in Eq. (44)) in Appendix A). These considerations lead to

$$\hat{B}(\hat{\pi}) = \frac{\hat{\omega}\hat{\psi}\left[1 - (1+\hat{\alpha}^2)\hat{\pi}\right]\hat{V}(\tilde{\mathbf{R}})}{1 - \hat{\omega}\hat{\psi}(1+\hat{\alpha}^2)\hat{V}(\tilde{\mathbf{R}})} \qquad (17)$$

It is noted that bias correction during the practical evaluation of bias expressions is ignored in Ref. [4], hence, Eq. (17) constitutes an improved calculation method, which is applied to an improved formula. Replacing the true values by bias-corrected quantities in Eq. (12) yields

$$\hat{B}(\hat{c}_u) = \frac{\hat{\alpha}\tilde{c}_0\hat{B}(\hat{\pi})}{(1-\hat{\pi})\left(1 - \hat{\pi} + \hat{B}(\hat{\pi})\right)} \qquad (18)$$

and one observes that evaluating Eq. (12) without bias correction would have led to a minus sign for $\hat{B}(\hat{\pi})$ in the denominator. Depending on the relative size of $\hat{B}(\hat{\pi})$, the difference can be noticeable.

In contrast, the estimated standard errors in $\hat{\pi}$ and $\hat{c}_u$ simply follow as

$$\hat{\sigma}(\hat{\pi}) = \hat{V}(\hat{\pi})^{1/2} = (\hat{r}_Q^*)^{-1}\left[(1-\hat{\pi})^2 + \hat{\alpha}^2\hat{\pi}^2\right]^{1/2}\hat{\sigma}(\tilde{\mathbf{R}}) \quad (19)$$

and

$$\hat{\sigma}(\hat{c}_u) = \hat{V}(\hat{c}_u)^{1/2} = \left[\hat{s}^{-2}\left(1+\hat{h}_u\right)\hat{V}(\tilde{\mathbf{R}}) + \hat{h}_u \cdot \hat{V}(\tilde{c}_0)\right]^{1/2} \quad (20)$$

where $\sigma(\tilde{\mathbf{R}}) = V(\tilde{\mathbf{R}})^{1/2}$. Unlike Eq. (15), Eq. (20) implicitly depends on the weight because a sensitivity and leverage are inserted that carry a weight-dependent bias.

3.6. Bias-corrected predictions

A bias correction amounts to subtracting the approximate bias from the biased quantity, i.e.,

$$\hat{c}_u^{bc} = \hat{c}_u - \hat{B}(\hat{c}_u) = \frac{\hat{\alpha}\tilde{c}_0\left(\hat{\pi} - \hat{B}(\hat{\pi})\right)}{\left(1-\hat{\pi}+\hat{B}(\hat{\pi})\right)} \quad (21)$$

Since the subtracted bias carries an uncertainty, this must be accounted for by an appropriately increased variance of the bias-corrected prediction. To first order, the variance in the bias-corrected prediction is approximated by

$$V\left(\hat{c}_u^{bc}\right) \approx \left(\frac{\partial c_u^{bc}}{\partial \alpha}\right)^2 V(\hat{\alpha}) + \left(\frac{\partial c_u^{bc}}{\partial c_0}\right)^2 V(\tilde{c}_0) + \left(\frac{\partial c_u^{bc}}{\partial \pi}\right)^2 V(\hat{\pi}) + \left(\frac{\partial c_u^{bc}}{\partial B(\hat{\pi})}\right)^2 V\left(\hat{B}(\hat{\pi})\right) \quad (22)$$

where the partial derivatives $(\partial c_u^{bc}/\partial \cdot)$ are evaluated at the errorless data ($\mathbf{R}_0$, $\mathbf{R}_u$ and c_0). The first term on the right-hand side of Eq. (22) is zero because, with errorless data matrices, a variation of α is cancelled by the resulting variation in π ($\alpha \neq 0$ by assumption). The remaining terms can be evaluated using

$$\frac{\partial c_u^{bc}}{\partial c_0} = g^{-1}\alpha\left(\pi - B(\hat{\pi})\right)$$

$$\frac{\partial c_u^{bc}}{\partial \pi} = g^{-1}\left(\alpha c_0 + c_u^{bc}\right)$$

$$\frac{\partial c_u^{bc}}{\partial B(\hat{\pi})} = -g^{-1}\left(\alpha c_0 + c_u^{bc}\right)$$

where $g = 1 - \pi + B(\hat{\pi})$.

The variance in the analyte concentration of the calibration sample, $\tilde{c}_0$, can be obtained by replication, the variance in the eigenvalue, $\hat{\pi}$, is given by Eq. (14), and the variance in the bias in the eigenvalue, $\hat{B}(\hat{\pi})$, follows from Eq. (64). Since Eq. (22) yields a variance, no bias-correction is applied to the inserted values (see Section 3.5).

3.7. Weights for the data matrices

In the original formulation of GRAM [1], both data matrices receive an equal weight, namely, $\alpha = 1$ (fixed). However, the bias correction with $\alpha = 1$ may be inefficient when bias dominates variance. The reason for this is that subtracting a large bias introduces a large uncertainty. A potential improvement is derived as follows. Eq. (10) reveals that the weight α has a profound influence on the bias in $\hat{\pi}$. In particular, approximate bias disappears altogether when

$$1 + \alpha^2 = \pi^{-1} = 1 + \alpha c_0 / c_u \quad (23)$$

from which it is immediate that the ideal value for α is given by

$$\alpha_{opt} = c_0 / c_u \quad (24)$$

Clearly, eliminating bias from $\hat{\pi}$ by direct application of Eq. (24) is impossible because c_u is unknown. This implies that an iterative procedure should be used to estimate α_{opt}: Start with $\hat{\alpha}^{(0)} = 1$, which leads to the 'ordinary' GRAM eigenvalue problem. This yields the prediction $\hat{c}_u^{(0)} = \tilde{c}_0\hat{\pi}^{(0)}/(1 - \hat{\pi}^{(0)})$, from which the first update of the weight is calculated as $\hat{\alpha}^{(1)} = \tilde{c}_0/\hat{c}_u^{(0)}$. Performing a new

GRAM calculation with $\hat{\alpha}^{(1)}$ gives the updated prediction $\hat{c}_u^{(1)}$, and so on. At convergence, one has estimates for α_{opt}, π and c_u that are related as

$$\hat{\alpha} = \tilde{c}_0/\hat{c}_u$$

$$\hat{\pi} = \hat{c}_u/\left(\hat{c}_u + \hat{\alpha}\tilde{c}_0\right)$$

$$\hat{c}_u = \hat{\alpha}\tilde{c}_0\hat{\pi}/(1-\hat{\pi}) \qquad (25)$$

The method where the weight α is estimated to reduce the bias in $\hat{\pi}$ will be referred to as iteratively reweighted GRAM (IRGRAM). It has to be repeated for each analyte.

The following remarks concerning the ideal weight, α_{opt}, seem to be in order. First, it is easily verified that a GRAM prediction based on $\mathbf{Q} = \mathbf{R}_u + \alpha_{opt}\mathbf{R}_0$ and $\mathbf{R}_u$ is equivalent to a GRAM prediction based on $\mathbf{Q} = c_u\mathbf{R}_u + c_0\mathbf{R}_0$ and $c_u\mathbf{R}_u$. In other words, this weight amounts to weighing each data matrix according to the analyte contribution to the data, which is attractive from a symmetry point of view. Second, to the authors' best knowledge this weighing procedure has no analogue in the univariate and multivariate domain.[2] In this paper we use Monte Carlo simulations to demonstrate its viability. Third, the ideal weight requires that $c_u \neq 0$. When predicting a strictly positive quantity such as analyte concentration, this condition reduces to $c_u > 0$. Consequently, application of IRGRAM becomes problematic when the analyte is present at trace levels. We are currently investigating alternatives to deal with this problem. Fourth, it is noted that, at convergence, the estimated bias in $\hat{\pi}$ will be zero (exactly). However, the true bias in $\hat{\pi}$ will not be zero because Eq. (10) is approximate and the 'ideal' weight has been determined from the predicted analyte concentration, which is certainly not free from error. It is seen that, in essence, IRGRAM amounts to subtracting an *uncertain zero* from $\hat{\pi}$. This uncertainty is effectively accounted for in the variance of the bias-corrected prediction through the last term in Eq. (22). Fifth, in certain applications, such as quality control, the same calibration sample is used several times. Under these circumstances, it could be worthwhile in terms of reduced prediction variance to measure the calibration data matrix to a higher precision than intended for the unknown samples. It is easily verified that the expression for the ideal weight changes to $\alpha_{opt} = (c_0/c_u)/(V(\tilde{\mathbf{R}}_0)/V(\tilde{\mathbf{R}}_u))$.

[2] For example, the seemingly related method of locally weighted regression has an entirely different background. However, Linder and Sundberg [16] have developed a method for the calibration of bilinear data where a similar weighing plays a key role. Interestingly, the weights are based on considerations about prediction variance; they follow from solving a least-squares problem. No weight is applied to the unknown sample, because it is not included in the modeling step. For more details, see Ref. [16].

3.8. Multiple calibration samples

GRAM is often discussed as a method that can only exploit a single calibration sample. However, a straightforward generalisation to multiple calibration samples is obtained by adding the calibration sample data matrices to emulate the single calibration sample case [13]. Thus, similar to Eq. (5), the relevant matrices are

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{R}}_u + \alpha\sum_{n=1}^{N}\tilde{\mathbf{R}}_n = \mathbf{H}\mathbf{Y}^T + \mathbf{E}_Q$$

$$\tilde{\mathbf{R}}_u = \mathbf{H}\mathbf{\Pi}\,\mathbf{Y}^T + \mathbf{E}_u \qquad (26)$$

where N denotes the number of calibration samples, $\mathbf{R}_n$ is the data matrix for the nth calibration sample, $\mathbf{H} = \mathbf{X}\left(\mathbf{C}_u + \alpha\sum_{n=1}^{N}\mathbf{C}_n\right)$ in which $\mathbf{C}_n$ contains the concentration for the nth calibration sample, and $\mathbf{\Pi} = \left(\mathbf{C}_u + \alpha\sum_{n=1}^{N}\mathbf{C}_n\right)^{-1}\mathbf{C}_u$.

All main results derived for the single-calibration-sample case carry over to the multiple-calibration-samples case after making the following substitutions:

- Eq. (10): insert the appropriate expressions for ψ and π.
- Eq. (12): replace c_o by the average analyte concentration for the calibration set, $\bar{c}$, and insert the adapted Eq. (10) and the appropriate expression for π.
- Eq. (14): insert the appropriate expressions for r_Q^* and π.
- Eq. (15): replace c_o^2 by $N\bar{c}^2$ in the expression for the leverage. It is noted that this substitution is at variance with the formula given in Ref. [8].
- Eq. (24): the ideal value for α is given by $\alpha_{opt} = \bar{c}/c_u$.

The effect of increasing the calibration set on α_{opt} is immediate, whereas the consequences for Eqs. (10), (12), and (14) are less clear. Straightforward manipulations show that the approximate prediction variance Eq. (15) will decrease if the analyte concentration in the additional sample exceeds the value $[(N^2 + N)^{1/2} - N]\bar{c}$. For $N = 1$, one obtains the value $(\sqrt{2} - 1)c_0 \approx 0.4c_0$, while for large N, this expression converges to $1/2\bar{c}$. These relationships should be useful for the design of calibration samples.

4. Experimental

Monte Carlo simulations are performed to test the adequacy of the proposed handling of prediction bias. These simulations amount to generating a series of independent data sets with artificially added noise. Bias can be determined from the deviation of the mean from the true value, while variance follows from the spread around the mean. The large number of replicates used in this work, namely 10000, ensures that the simulation results are much more precise than the formula-based quantities.

4.1. Construction of the data matrices

Three-component systems are simulated by multiplying Gaussian elution profiles (Table 1) by experimentally obtained ultraviolet (UV) spectra of adenine, cytidine and guanine [17] (Fig. 3). Thus, the column profiles are elution profiles while the row profiles constitute spectra. In the remainder of the paper, it should be clear from the context what is meant by the term 'profile'. The spectra are normalised to make the contribution to the total variance proportional to the square of the peak height. Normally distributed noise is added with the standard deviation $\sigma(\tilde{\mathbf{R}})$ taking the values 0.5, 1 and 2, respectively. The number of spectra or time steps (J_1) is 800, while the number of elution profiles or UV channels (J_2) is 36.

Table 1
Description of simulated elution profiles

Simulation parameter	Adenine	Cytidine	Guanine
Peak position	300	400	500
Standard deviation peak	120	120	120
Peak height calibration sample	100	100	100
Peak height unknown sample	200	100	10

4.2. Variances of the measurement errors

The practical evaluation of expressions derived in this paper requires estimates for $V(\tilde{\mathbf{R}})$ and $V(\tilde{c}_0)$. It is noted that for all calculations, $V(\tilde{\mathbf{R}})$ is kept fixed while $V(\tilde{c}_0) = 0$. Maintaining a fixed value for $V(\tilde{\mathbf{R}})$ is based on the consideration that the formula-based quantities are validated with the results of extensive Monte Carlo simulations. Stated differently: the purpose of the Monte Carlo simulations is to verify that the predicted amount of error propagation is correct, given the size (variance) of the error. As long as both approaches (formula and Monte Carlo simulations) use the same value for $V(\tilde{\mathbf{R}})$, it is immaterial whether this variance is estimated or fixed.

Often, good noise estimates can be obtained using replication [18]. However, Wang and Hopke [19] have shown that one can also estimate the noise from the residuals of SVD. The underlying assumption of their method is that the data sets are large enough so that the standard deviation of the noise is constant over a small region. Then, averaging the squared residuals of a certain number of neighbours yields a variance estimate for the data element under consideration. Although their method only leads to estimates of heteroscedasticity, it is easily adapted to account for correlations by also averaging cross products of the residuals. Formally, correct application of this method requires that the dimensionality of the SVD model be known, e.g., from prior knowledge about the constitution of the sample.

The condition $V(\tilde{c}_0) = 0$ amounts to not investigating the effect of the uncertainty in the concentration in the calibration sample. The reason for doing so is that this uncertainty invariably contributes through terms that are not specific for GRAM. (It is immaterial which method yields the concentration ratio, which now follows from solving an eigenvalue problem.) Moreover, the contribution of $V(\tilde{c}_0)$ be-

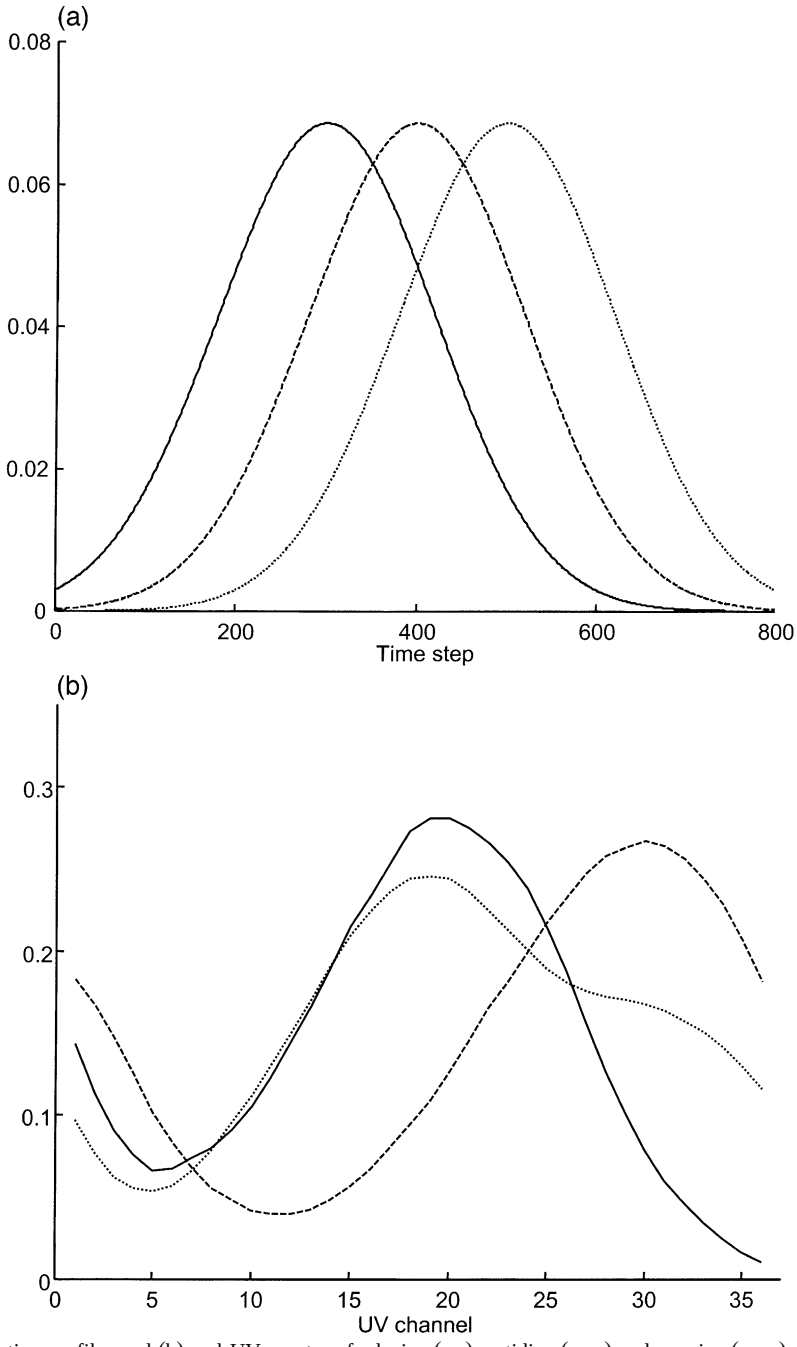

Fig. 3. (a) Simulated elution profiles and (b) real UV spectra of adenine (—), cytidine (- - -) and guanine (· · ·). The elution profiles are normalised to remove sample-dependency through peak heights.

comes apparent through manipulating scalar expressions. It is reasonable to assume that the approximations involved need not be validated, since they are fairly standard.

4.3. Calculations

All calculations were performed in Matlab version 4.2c (The Mathworks, Natick, MA). A copy of the program is available on request.

5. Results and discussion

The simulations are designed to produce the largest relative bias for guanine; hence, most attention will be directed to this analyte. It is further noted that, when using IRGRAM for optimising the weight factor for guanine, the results may deteriorate for adenine and cytidine. This effect is not shown, since it is implied by the procedure. (IRGRAM should be applied to each analyte separately.)

5.1. Description of the data matrices

Inspection of variance and bias expressions shows that the quality of the data is determined by the combination of overlap and level of the noise; hence, the individual characterisation of elution profiles (Table 1) and spectra (Fig. 3) only yields limited insight. Useful summary statistics are the analytical figures of merit proposed by Faber et al. [8] (Table 2). The NAS specifies the amount of the pure analyte data matrix that contributes to the model. According to Eq. (14), it determines how instrumental errors propagate to the variance in the eigenvalues. The sensitivity, being the ratio of NAS and analyte concentrations, can be compared with the slope of a univariate calibration graph [8]. According to Eq. (15), it determines how instrumental errors propagate to the variance in the predictions. For convenience, peak height plays the role of analyte concentration in this work. (Thus, the sensitivity is obtained as the ratio of the NAS, given in Table 2, and the appropriate peak height given in Table 1.) Finally, the selectivity gives the fraction of the pure analyte data matrix that contributes to the model. It is observed that approximately 90% of the pure analyte data matrices does not contribute to the model because of overlap with the interferences. The selectivity values can be used to assess to what extent propagation of instrumental errors can be reduced, which is important information if a method is to be designed or optimised to meet specific needs.

Table 2
Analytical figures of merit for data matrices

Figure of merit	Adenine	Cytidine	Guanine
NAS calibration sample	148	156	112
NAS unknown sample	296	156	11.2
Sensitivity	1.48	1.56	1.12
Selectivity	0.10	0.11	0.08

5.2. Bias in the estimated profiles

It is derived in Appendix A that bias in the profile estimates does not contribute to bias in the predictions, which is somewhat counterintuitive. The validity of this result is, however, easily verified by inserting bias-corrected profiles in Eq. (17). Since we have not been able to derive an expression for this bias, we resort to the simulation method detailed in Appendix D. Using the setting $N_{\text{add}} = 1000$, $V_1 = 0.1$ and $V_2 = 0.2$, the largest effect of the bias correction is observed at the intermediate noise level, i.e., $\sigma(\tilde{\mathbf{R}}) = 1$ and weight 10 (or 9.76): the RMSE is reduced by a factor 3 (Fig. 4). Even for this rather extreme case, inserting the bias-corrected profile estimates in Eq. (17) did not have a palpable effect.

5.3. Bias in the eigenvalues if ideal weight is used

IRGRAM is based on the assumption that bias vanishes if the ideal weight is used. This assumption is easily verified by applying the standard GRAM to cytidine, since the true cytidine peak heights are identical for the calibration and unknown sample. The Monte Carlo simulations show that prediction bias is orders of magnitude smaller than the standard error; the RMSE is hardly affected by prediction bias (Table 3). The standard error in the estimated bias is calculated as the standard error in the mean from which the bias is calculated. Since the mean is based on 10,000 replications, this standard error is 1% of the standard error in a single replication. (The latter stan-

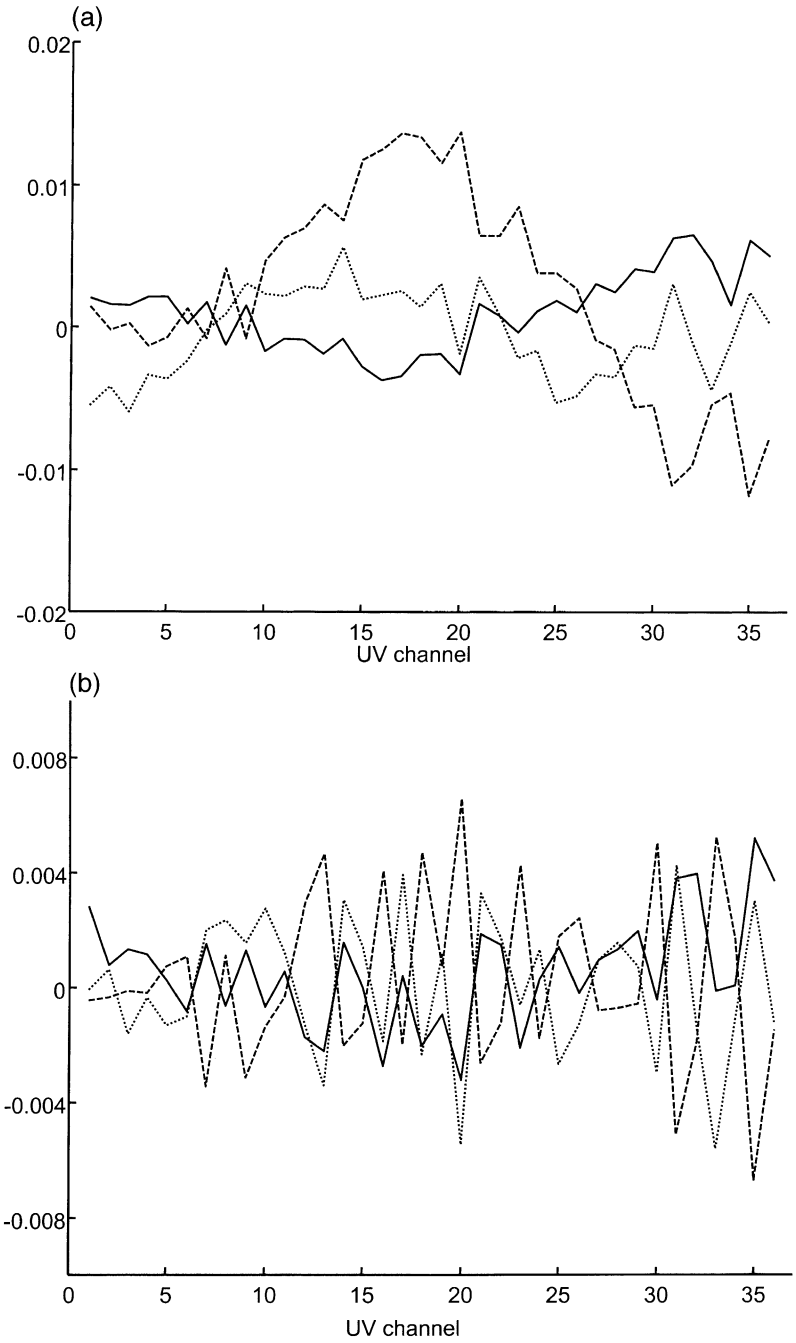

Fig. 4. Errors in estimated UV spectra of adenine (—), cytidine (- - -) and guanine (· · ·): (a) without bias correction, and (b) after bias correction.

dard error is reported in the next column.) Taking into account that these simulations are designed to produce a large prediction bias (see below), it can be inferred that the cancellation in Eq. (10) is effective.

Table 3
Monte Carlo estimates of errors (bias, standard error and RMSE) in estimated eigenvalues for cytidine ($\alpha = \alpha_{opt} = 1$). The numbers in parentheses denote the standard error in the last reported digits. The symbols are explained in the text

$\sigma(\tilde{\mathbf{R}})$	$\hat{B}(\hat{\pi})$ (%)	$\hat{\sigma}(\hat{\pi})$ (%)	$\hat{RMSE}(\hat{\pi})$ (%)[a]
0.5	+0.002 (2)	0.23	0.23
1	+0.008 (5)	0.46	0.46
2	+0.029 (9)	0.94	0.94

[a]Calculated using Eq. (3).

5.4. Bias in the eigenvalues as a function of the selected weight

The results for guanine are summarised in Table 4. The second column gives the value for the weight factor. Since the eigenvalue itself varies with the weight, all uncertainties (bias, standard error and RMSE) are reported as a percentage of the true value to enhance the interpretability. Selecting the weight unity, which amounts to the standard GRAM calculation, yields a large bias in the eigenvalues. A weight of 10 is ideal for guanine, since the peak heights are 100 and 10 for the calibration and unknown samples, respectively. The values 9.96, 9.76 and 9.01 are the estimates produced by IRGRAM. (Convergence is obtained in 5, 7 and 11 iterations, respectively, when iterating until the relative change in the weight falls below 10^{-6}.) As expected, the deviation from the ideal value of 10 tends to increase with increasing noise level. It is important to note, however, that the weight 9.01 has led to results that are quite similar to the ones obtained using the ideal weight. This observation suggests that the results are not very sensitive with respect to the value of the weight, which lends support to IRGRAM.

Using the ideal weight of 10 leads to a small bias at all noise levels: the Monte Carlo estimates are hardly larger than the associated standard error. Estimating the ideal weight yields a zero bias estimate, which is indicated by the associated standard errors. It seems that these standard errors are less influenced by the exact value of the weight than the bias estimates themselves.

The old calculation method for bias [4] yields the values +11.1%, +29.8% and +49% for the standard GRAM (no standard error was derived in Ref. [4]), instead of +12.8(9)%, +45(6)% and +108(26)%, when using Eqs. (17) and (64). The improvement is obvious from a comparison with the Monte Carlo results: +12.94(4)%, +47.26(7)% and +140.0(1)%. Only for the highest noise level, the estimate obtained from Eq. (17) is somewhat too small. The excellent agreement between the formula-based bia and the Monte Carlo values as the lowest two noise levels suggests that adequate bias correction is feasible during the practical evaluation of bias expressions.

Table 4
Estimated eigenvalues and associated error estimates (bias, standard error and RMSE) for guanine ($\alpha_{opt} = 10$). The numbers in parentheses denote the standard error in the last reported digits. The symbols are explained in the text

$\sigma(\tilde{\mathbf{R}})$	$\hat{\alpha}$	π ($\times 10^{-2}$)	$\hat{\pi}$ ($\times 10^{-2}$)	Formula		Monte Carlo		$\hat{RMSE}(\hat{\pi})$(%)[a]
				$\hat{B}(\hat{\pi})$(%)[b]	$\hat{\sigma}(\hat{\pi})$(%)[c]	$\hat{B}(\hat{\pi})$(%)	$\hat{\sigma}(\hat{\pi})$(%)	
0.5	1	9.09	10.3	+12.8 (9)	3.95	+12.94 (4)	3.93	13.5
	10	0.99	0.99	−0.01 (7)	4.43	−0.05 (4)	4.42	4.42
	9.96	0.99	1.00	0.00 (7)	4.43	−0.04 (4)	4.42	4.42
1	1	9.09	13.6	+45 (6)	7.17	+47.26 (7)	7.05	47.8
	10	0.99	1.01	−0.2 (6)	8.71	−0.10 (9)	8.68	8.68
	9.76	1.01	1.04	0.0 (6)	8.72	+0.05 (9)	8.68	8.68
2	1	9.09	22.5	+108 (26)	10.5	+140.0 (1)	10.4	140
	10	0.99	1.07	−2 (4)	16.3	−0.2 (2)	16.2	16.2
	9.01	1.10	1.22	0 (4)	16.3	+2.2 (2)	16.2	16.3

[a]Calculated using Eq. (3).
[b]Calculated using Eq. (17).
[c]Calculated using Eq. (19).

Close inspection of the Monte Carlo results shows that the bias does not scale up as expected from Eq. (10): doubling the size of the standard deviation of the measurement errors does not quadruple the size of the bias. The deviation from the 'ideal' behaviour is only moderate for the intermediate noise level, which explains why the bias expression still works well in that regime. However, this approximate formula must break down when going to the highest noise level, since bias scales up by a factor of three instead of four. It follows that the formula-based standard error cannot be trusted either, although it appears to be reasonable.

The bias addition method ($N_{add} = 1000$, $V_1 = 0.1$ and $V_2 = 0.2$) yields the values 12.2(5)%, 40(2)% and 84(5)%. Thus, only at the lowest noise level, this method works well with the current setting. Moreover, the failure at the higher noise levels is not reflected in the standard error obtained as the square root of Eq. (78), which is somewhat disturbing. This failure can, however, be conveniently explained. When adding noise at two levels, one assumes that the bias increases linear with variance. For the current data sets, which are especially constructed to lead to large biases, this assumption is not realistic. It seems likely that the results can be improved by adding noise at more than two levels. However, this possibility is not pursued here, since the noise addition method is merely used to confirm that bias in the profile estimates does not contribute to bias in the predictions.

5.5. Bias in the predictions as a function of the selected weight

In Table 5, the uncertainties in the prediction (bias, standard error and RMSE) are reported as a percentage of the true values to facilitate the comparison with the numbers in Table 4. The prediction bias follows similar trends as the bias in the eigenvalues.

5.6. Standard error in the predictions as a function of the selected weight

The formula-based standard error in the predictions varies slightly with the applied weight (column 6 of Table 5). This behaviour is not expected from Eq. (15), which, unlike the bias expression (12), does not contain the weight. However, a slight dependence on the weight is expected from the practical evaluation of Eq. (15) using Eq. (20). The reason for this is that the inserted values for sensitivity and leverage are biased. The excellent agreement between the formula-based standard errors and the Monte Carlo values corroborates that biased values should be inserted in Eq. (20). Since the leverage term only makes a tiny contribution to prediction variance, the remainder of the discussion focuses on sensitivity. The

Table 5
Estimated peak heights and associated error estimates (bias, standard error and RMSE) for guanine ($\alpha_{opt} = 10$). The numbers in parentheses denote the standard error in the last reported digits. The symbols are explained in the text

$\sigma(\tilde{\mathbf{R}})$	$\hat{\alpha}$	$\hat{H}$	$\hat{s}$	Formula		Monte Carlo		$\hat{RMSE}(\hat{H})$(%)[a]
				$\hat{B}(\hat{H})$(%)[b]	$\hat{\sigma}(\hat{H})$(%)[c]	$\hat{B}(\hat{H})$(%)	$\hat{\sigma}(\hat{H})$(%)	
0.5	1	11.48	1.13	+14.3 (10)	4.46	+14.44 (4)	4.43	15.1
	10	10.04	1.12	−0.01 (8)	4.47	−0.05 (4)	4.46	4.46
	9.96	10.04	1.12	0.00 (8)	4.47	−0.04 (4)	4.46	4.46
1	1	15.71	1.16	+52 (7)	8.73	+54.63 (9)	8.55	55.3
	10	10.23	1.14	−0.2 (6)	8.80	−0.10 (9)	8.77	8.77
	9.76	10.24	1.14	0.0 (6)	8.81	+0.06 (9)	8.77	8.77
2	1	28.96	1.31	+145 (32)	15.8	+179.2 (2)	15.4	180
	10	10.86	1.22	−2 (4)	16.5	−0.2 (2)	16.4	16.4
	9.01	11.10	1.22	0 (4)	16.5	+2.3 (2)	16.3	16.5

[a] Calculated using Eq. (3).
[b] Calculated using Eq. (18).
[c] Calculated using Eq. (20).

bias in the sensitivity becomes clear from the values reported in column 4 of Table 5: sensitivity is biased upwards for all cases considered in this work (the true value is 1.12, see Table 2). As a consequence, the propagation of measurement errors is smaller than expected from the true value 1.12. In addition, it is observed that the sensitivity is largest for the cases with unity weight. Deriving a bias expression for the sensitivity could lead to a better understanding of these effects. This question is, however, not further pursued here.

5.7. Adequacy of bias correction

The predicted peak heights in column 3 of Table 5 are bias-corrected using the bias estimates given in column 5 of Table 5 (see Table 6). The standard errors in the bias-corrected peak heights are given as absolute values now because these numbers constitute the final results. As expected from the results presented above, applying IRGRAM is almost equivalent to fixing the weight to the ideal value (10) at all noise levels. At the lowest noise level, the original GRAM performs as good as IRGRAM because the subtracted bias carries an uncertainty that is negligible compared with the standard error in the original quantity (0.10% in column 5 versus 4.46% in column 6). At the intermediate noise level, however, the original GRAM leads to a 20% larger standard error in the bias-corrected prediction because a bias is subtracted that carries a relatively large uncertainty (7% versus 8.73%). Obviously, the difference between the original GRAM and IRGRAM is most pronounced at the highest noise level: the standard error in the bias-corrected prediction is smaller by a factor of two by applying the iteratively determined weight first (see Fig. 2).

Table 6
Results of bias correction. The symbols are explained in the text

$\sigma(\tilde{\mathbf{R}})$	$\hat{\alpha}$	$\hat{H}^{\mathrm{bca}}$	$\hat{\sigma}(\hat{H}^{\mathrm{bc}})^{\mathrm{b}}$
0.5	1	10.06	0.45
	10	10.04	0.45
	9.96	10.04	0.45
1.0	1	10.49	1.06
	10	10.24	0.88
	9.76	10.24	0.88
2.0	1	14.46	3.39
	10	11.09	1.71
	9.01	11.10	1.71

[a] Calculated using Eq. (21).
[b] Calculated using Eq. (22).

6. Conclusions and outlook

Substantial progress has been achieved with respect to the handling of prediction bias when using GRAM. In the current simulation study, IRGRAM has performed best in the sense that the increase of prediction variance due to subtracting bias is smallest. The results suggest that the exact value of the weight is not critical, which is important from the practical point of view. Future research should be concerned with testing the developed methodology in practical situations where heteroscedastic and correlated noise may be encountered (see Appendix C). It is expected that the improved handling of prediction bias will impact method comparison studies. It should be interesting to see, for example, how the improved GRAM compares with a competing method such as alternating least squares (ALS).

The problem of determining bias in the profile estimates has been approached using a simulation technique where noise is added to the data matrices. However, obtaining satisfactory results using noise additions implies a considerable increase of workload; hence, it is desirable that bias expressions be derived for the profile estimates as well.

It is remarkable that the prediction variance depends only slightly on the weighing of the data; the weight factor α primarily affects prediction bias. There is a similarity between the currently developed bias correction device and the method of principal covariate regression (PCovR) [20] because PCovR also contains a weighing factor that is optimised with respect to prediction error. In PCovR, an eigenvalue problem is solved to calculate scores that yield an optimal reconstruction of both the predictor matrix $\mathbf{X}$ and the predictand matrix $\mathbf{Y}$. Investigating how the weight factor in PCovR affects prediction variance and bias could be an interesting subject for future research.

Appendix A. Derivation of Eq. (10)

Deriving an accurate expression for bias in the eigenvalues is complicated. In fact, the purpose of the current derivation is to replace an earlier attempt [4] that was only partly successful. To better understand the basic strategy, it is illustrative to look at a univariate example first. Consider the problem of estimating the quotient of two random variables, i.e., $\hat{z} = \tilde{x}/\tilde{y}$. This example is taken from the work of Moran and Kowalski [21] who examined the effect of random experimental error on variance and bias in the initial concentrations calculated by the generalized standard addition method (GSAM). No assumption is made about the shape of the distribution of $\tilde{x}$ and $\tilde{y}$ (e.g., normal or uniform). It is only required that the variances be finite, the input data should be unbiased, i.e., $E(\tilde{x}) = x$ and $E(\tilde{y}) = y$, and $y \neq 0$ (to avoid a possible divide by zero). Expanding $\hat{z}$ to second order yields

$$\begin{aligned}\hat{z} &= (x + \mathrm{d}x)/(y + \mathrm{d}y)\\ &= (x + \mathrm{d}x)\left(y^{-1} - y^{-2}\mathrm{d}y + y^{-3}(\mathrm{d}y)^2 - \ldots\right)\\ &\approx z + \mathrm{d}z + \mathrm{d}^2 z\\ &= (x + \mathrm{d}x)\left(y^{-1} - y^{-2}\mathrm{d}y + y^{-3}(\mathrm{d}y)^2\right) \qquad (27)\end{aligned}$$

This expansion holds provided the relative error $y^{-1}\mathrm{d}y$ is less than unity, which is a reasonable requirement. The bias in $\hat{z}$ is defined as $B(\hat{z}) \equiv E(\hat{z}) - z$. Taking expectation of Eq. (27) shows that the first-order terms will not contribute because the data is unbiased by assumption. It follows that determining bias amounts to evaluating the second differential. Discarding terms of order higher than two gives

$$\begin{aligned}B(\hat{z}) &\approx E(\mathrm{d}^2 z)\\ &= z\left(y^{-2}E\left((\mathrm{d}y)^2\right) - x^{-1}y^{-1}E(\mathrm{d}x\mathrm{d}y)\right)\\ &= z\left(y^{-2}V(\tilde{y}) - x^{-1}y^{-1}C(\tilde{x},\tilde{y})\right) \qquad (28)\end{aligned}$$

It is observed that the approximate bias in $\hat{z}$ contains a variance as well as a covariance term that tend to cancel. This behaviour is also expected for the approximate bias in the eigenvalues because determining the desired concentration ratios can be interpreted as a 'division' of matrices for which the individual elements are correlated for the currently studied modification of GRAM.

The preceding discussion suggests that working out the appropriate expression for the second differential of the eigenvalue function [11] can yield an approximate bias. However, working out this expression requires making the errors in the diagonalised matrix explicit, which constitutes a formidable task, since this matrix is a product of four matrices of which three are estimated themselves. Thus, it is preferred to start from the expression that enabled the straightforward derivation of variance expressions [2,4], i.e.,

$$\hat{\mathbf{\Pi}} = \hat{\mathbf{H}}^{+}\tilde{\mathbf{R}}_{\mathrm{u}}\hat{\mathbf{Y}}^{+\mathrm{T}} \qquad (29)$$

To second order, $\hat{\mathbf{\Pi}}$ is approximated as

$$\hat{\mathbf{\Pi}} \approx \mathbf{\Pi} + \mathrm{d}\mathbf{\Pi} + \mathrm{d}^2\mathbf{\Pi} \qquad (30)$$

and the goal of the derivation is to determine the bias in $\hat{\mathbf{\Pi}}$ by working out

$$B(\hat{\mathbf{\Pi}}) \equiv E(\hat{\mathbf{\Pi}}) - \mathbf{\Pi} \approx E(\mathrm{d}^2\mathbf{\Pi}) \qquad (31)$$

First, we will introduce some notation that allows the straightforward expansion of Eq. (29). By combining Eqs. (6) and (8), $\tilde{\mathbf{Q}}$ is expressed as

$$\tilde{\mathbf{Q}} = \hat{\mathbf{H}}\hat{\mathbf{Y}}^{\mathrm{T}} + \hat{\mathbf{E}}_{\mathrm{Q}} \qquad (32)$$

so that

$$\begin{aligned}\hat{\mathbf{H}} &= \tilde{\mathbf{Q}}\hat{\mathbf{Y}}^{+\mathrm{T}}\\ \hat{\mathbf{Y}} &= \tilde{\mathbf{Q}}^{\mathrm{T}}\hat{\mathbf{H}}^{+\mathrm{T}} \qquad (33)\end{aligned}$$

To second order, these profile estimates are approximated as

$$\begin{aligned}\hat{\mathbf{H}} &\approx \mathbf{H} + \mathrm{d}\mathbf{H} + \mathrm{d}^2\mathbf{H}\\ \hat{\mathbf{Y}} &\approx \mathbf{Y} + \mathrm{d}\mathbf{Y} + \mathrm{d}^2\mathbf{Y} \qquad (34)\end{aligned}$$

The key step in the derivation is to construct a decomposition of $\tilde{\mathbf{R}}_u$ that is similar to Eq. (32), namely,

$$\tilde{\mathbf{R}}_u = \hat{\mathbf{H}}_u \boldsymbol{\Pi} \hat{\mathbf{Y}}_u^T + \hat{\mathbf{E}}_u \tag{35}$$

The requirement for the artificially constructed profiles $\hat{\mathbf{H}}_u$ and $\hat{\mathbf{Y}}_u$ is that they capture all correlation between the data matrices $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}_u$ so that, eventually, the analogue of Eq. (28) can be obtained. (In particular, there should be no contribution from the residual matrix $\hat{\mathbf{E}}_u$.) Some afterthought shows that $\hat{\mathbf{H}}_u$ and $\hat{\mathbf{Y}}_u$ are found, similar to $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$ in Eq. (33), as

$$\hat{\mathbf{H}}_u = \tilde{\mathbf{R}}_u \hat{\mathbf{Y}}_u^{+T} \boldsymbol{\Pi}^{-1}$$

$$\hat{\mathbf{Y}}_u = \tilde{\mathbf{R}}_u^T \hat{\mathbf{H}}_u^{+T} \boldsymbol{\Pi}^{-1} \tag{36}$$

The inverse eigenvalue matrix provides the correct scaling. It will cancel in the sequel so that a division by zero will not occur. Since the true eigenvalues are involved, $\hat{\mathbf{H}}_u$ and $\hat{\mathbf{Y}}_u$ cannot be evaluated. To second order, these artificial profiles are approximated as

$$\hat{\mathbf{H}}_u \approx \mathbf{H} + \mathrm{d}\mathbf{H}_u + \mathrm{d}^2\mathbf{H}_u$$

$$\hat{\mathbf{Y}}_u \approx \mathbf{Y} + \mathrm{d}\mathbf{Y}_u + \mathrm{d}^2\mathbf{Y}_u \tag{37}$$

Inserting Eq. (35) in Eq. (29) yields

$$\hat{\boldsymbol{\Pi}} = \hat{\mathbf{H}}^+ \left(\hat{\mathbf{H}}_u \boldsymbol{\Pi} \hat{\mathbf{Y}}_u^T + \hat{\mathbf{E}}_u \right) \hat{\mathbf{Y}}^{+T} \tag{38}$$

which is worked out by expanding the estimated pseudo-inverse matrices in terms of the differentials of $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$. By assumption, $\mathbf{H}$ and $\mathbf{Y}$ are full column rank matrices. Unless the estimation errors are so large that GRAM breaks down, this property carries over to $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$ so that

$$\hat{\mathbf{H}}^+ = \left(\hat{\mathbf{H}}^T \hat{\mathbf{H}} \right)^{-1} \hat{\mathbf{H}}^T$$

$$\hat{\mathbf{Y}}^+ = \left(\hat{\mathbf{Y}}^T \hat{\mathbf{Y}} \right)^{-1} \hat{\mathbf{Y}}^T \tag{39}$$

Using Eq. (34), the first cross-product matrix in Eq. (39) is approximated as

$$\hat{\mathbf{H}}^T \hat{\mathbf{H}} \approx \mathbf{H}^T\mathbf{H} + \mathbf{H}^T\mathrm{d}\mathbf{H} + (\mathrm{d}\mathbf{H}^T)\mathbf{H} + (\mathrm{d}\mathbf{H}^T)\mathrm{d}\mathbf{H} + \mathbf{H}^T\mathrm{d}^2\mathbf{H} + (\mathrm{d}^2\mathbf{H}^T)\mathbf{H} \tag{40}$$

It is important to note that Eq. (40) is different from the expression given by Hodges and Moore [22] in that the second differential of $\hat{\mathbf{H}}$ is included. The reason for this is that Hodges and Moore [22] are concerned with quantifying the biasing effect of measurement errors in the predictors when using OLS. Measurement errors are completely described with the first differential. In contrast, we are trying to determine the biasing effect of estimation errors. These errors should be approximated to second order.

Defining $\mathbf{B}_H = \mathbf{H}^T\mathbf{H}$ and $\mathbf{D}_H = \mathbf{H}^T\mathrm{d}\mathbf{H} + (\mathrm{d}\mathbf{H}^T)\mathbf{H} + (\mathrm{d}\mathbf{H}^T)\mathrm{d}\mathbf{H} + \mathbf{H}^T\mathrm{d}^2\mathbf{H} + (\mathrm{d}^2\mathbf{H}^T)\mathbf{H}$, one obtains that

$$\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\right)^{-1} = (\mathbf{B}_H + \mathbf{D}_H)^{-1} \approx \mathbf{B}_H^{-1} - \mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1} + \mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1} - \ldots \tag{41}$$

This expansion holds provided the eigenvalues of $\mathbf{B}_H^{-1}\mathbf{D}_H$ are less than unity in absolute value [22]. Likewise,

$$\left(\hat{\mathbf{Y}}^T\hat{\mathbf{Y}}\right)^{-1} = (\mathbf{B}_Y + \mathbf{D}_Y)^{-1} \approx \mathbf{B}_Y^{-1} - \mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1} + \mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1} - \ldots \tag{42}$$

where $\mathbf{B}_Y = \mathbf{Y}^T\mathbf{Y}$ and $\mathbf{D}_Y = \mathbf{Y}^T\mathrm{d}\mathbf{Y} + (\mathrm{d}\mathbf{Y}^T)\mathbf{Y} + (\mathrm{d}\mathbf{Y}^T)\mathrm{d}\mathbf{Y} + \mathbf{Y}^T\mathrm{d}^2\mathbf{Y} + (\mathrm{d}^2\mathbf{Y}^T)\mathbf{Y}$. Inserting Eqs. (34), (37), (39), (41) and (42) in Eq. (38) yields that

$$\begin{aligned}
\hat{\boldsymbol{\Pi}} &\approx \boldsymbol{\Pi} + \mathrm{d}\boldsymbol{\Pi} + \mathrm{d}^2\boldsymbol{\Pi} \\
&= \left(\mathbf{B}_H^{-1} - \mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1} + \mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1}\mathbf{D}_H\mathbf{B}_H^{-1}\right) \\
&\quad \times (\mathbf{H} + \mathrm{d}\mathbf{H} + \mathrm{d}^2\mathbf{H})^T \Big[\left(\mathbf{H} + \mathrm{d}\mathbf{H}_u + \mathrm{d}^2\mathbf{H}_u\right) \\
&\quad \times \boldsymbol{\Pi} \left(\mathbf{Y} + \mathrm{d}\mathbf{Y}_u + \mathrm{d}^2\mathbf{Y}_u\right)^T + \mathrm{d}\mathbf{R}_u \Big] \\
&\quad \times (\mathbf{Y} + \mathrm{d}\mathbf{Y} + \mathrm{d}^2\mathbf{Y}) \left(\mathbf{B}_Y^{-1} - \mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1} \right. \\
&\quad \left. + \mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1}\mathbf{D}_Y\mathbf{B}_Y^{-1}\right)
\end{aligned} \tag{43}$$

and equating second-order terms leads, after direct cancellation of a number of terms, to

$$
\begin{aligned}
\mathrm{d}^2\mathbf{\Pi} \approx\ & \left\{\mathbf{B}_\mathrm{H}^{-1}\mathrm{d}\mathbf{H}^\mathrm{T}(\mathrm{d}\mathbf{H}_\mathrm{u})\mathbf{\Pi} - \mathbf{B}_\mathrm{H}^{-1}\mathrm{d}\mathbf{H}^\mathrm{T}(\mathrm{d}\mathbf{H})\mathbf{\Pi}\right.\\
&+\mathbf{B}_\mathrm{H}^{-1}(\mathrm{d}\mathbf{H}^\mathrm{T})\mathbf{P}_\mathrm{H}(\mathrm{d}\mathbf{H})\mathbf{\Pi}\\
&-\mathbf{B}_\mathrm{H}^{-1}(\mathrm{d}\mathbf{H}^\mathrm{T})\mathbf{P}_\mathrm{H}(\mathrm{d}\mathbf{H}_\mathrm{u})\mathbf{\Pi}\\
&+\mathbf{H}^{+}(\mathrm{d}\mathbf{H})\mathbf{H}^{+}(\mathrm{d}\mathbf{H})\mathbf{\Pi}\\
&-\mathbf{H}^{+}(\mathrm{d}\mathbf{H})\mathbf{H}^{+}(\mathrm{d}\mathbf{H}_\mathrm{u})\mathbf{\Pi}\\
&+\mathbf{\Pi}\mathrm{d}\mathbf{Y}_\mathrm{u}^\mathrm{T}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1} - \mathbf{\Pi}\mathrm{d}\mathbf{Y}^\mathrm{T}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1}\\
&+\mathbf{\Pi}(\mathrm{d}\mathbf{Y}^\mathrm{T})\mathbf{P}_\mathrm{Y}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1}\\
&-\mathbf{\Pi}\left(\mathrm{d}\mathbf{Y}_\mathrm{u}^\mathrm{T}\right)\mathbf{P}_\mathrm{Y}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1}\\
&+\mathbf{\Pi}(\mathrm{d}\mathbf{Y}^\mathrm{T})\mathbf{Y}^{+\mathrm{T}}(\mathrm{d}\mathbf{Y}^\mathrm{T})\mathbf{Y}^{+\mathrm{T}}\\
&\left.-\mathbf{\Pi}\left(\mathrm{d}\mathbf{Y}_\mathrm{u}^\mathrm{T}\right)\mathbf{Y}^{+\mathrm{T}}(\mathrm{d}\mathbf{Y}^\mathrm{T})\mathbf{Y}^{+}\right\}^\mathrm{T}\\
&+\left\{\mathbf{B}_\mathrm{H}^{-1}(\mathrm{d}\mathbf{H}^\mathrm{T})\mathbf{H}\mathbf{\Pi}\left(\mathrm{d}\mathbf{Y}_\mathrm{u}^\mathrm{T}\right)\mathbf{Y}^{+\mathrm{T}}\right.\\
&+\mathbf{B}_\mathrm{H}^{-1}(\mathrm{d}\mathbf{H}^\mathrm{T})\mathbf{R}_\mathrm{u}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1}\\
&-\mathbf{B}_\mathrm{H}^{-1}(\mathrm{d}\mathbf{H}^\mathrm{T})\mathbf{H}\mathbf{\Pi}\mathbf{D}_\mathrm{Y}\mathbf{B}_\mathrm{Y}^{-1}\\
&+\mathbf{H}^{+}(\mathrm{d}\mathbf{H}_\mathrm{u})\mathbf{\Pi}\,\mathbf{Y}^\mathrm{T}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1} - \mathbf{H}^{+}(\mathrm{d}\mathbf{H}_\mathrm{u})\\
&\times\mathbf{\Pi}\mathbf{D}_\mathrm{Y}\mathbf{B}_\mathrm{Y}^{-1} - \mathbf{B}_\mathrm{H}^{-1}\mathbf{D}_\mathrm{H}\mathbf{\Pi}\left(\mathrm{d}\mathbf{Y}_\mathrm{u}^\mathrm{T}\right)\mathbf{Y}^{+\mathrm{T}}\\
&-\mathbf{B}_\mathrm{H}^{-1}\mathbf{D}_\mathrm{H}\mathbf{\Pi}\,\mathbf{Y}^\mathrm{T}(\mathrm{d}\mathbf{Y})\mathbf{B}_\mathrm{Y}^{-1}\\
&\left.+\mathbf{B}_\mathrm{H}^{-1}\mathbf{D}_\mathrm{H}\mathbf{\Pi}\mathbf{D}_\mathrm{Y}\mathbf{B}_\mathrm{Y}^{-1}\right\}\\
&+\left\{-\mathbf{H}^{+}(\mathrm{d}^2\mathbf{H})\mathbf{\Pi} - \mathbf{\Pi}(\mathrm{d}^2\mathbf{Y}^\mathrm{T})\mathbf{Y}^{+\mathrm{T}}\right.\\
&\left.+\mathbf{H}^{+}\left(\mathrm{d}^2\mathbf{H}_\mathrm{u}\right)\mathbf{\Pi} + \mathbf{\Pi}\left(\mathrm{d}^2\mathbf{Y}_\mathrm{u}^\mathrm{T}\right)\mathbf{Y}^{+>\mathrm{T}}\right\}\\
=\ & A + B + C
\end{aligned}
\tag{44}
$$

where $\mathbf{P}_\mathrm{H} = \mathbf{H}\mathbf{B}_\mathrm{H}^{-1}\mathbf{H}^\mathrm{T}$ and $\mathbf{P}_\mathrm{Y} = \mathbf{Y}\mathbf{B}_\mathrm{Y}^{-1}\mathbf{Y}^\mathrm{T}$ project onto the column space of $\mathbf{H}$ and $\mathbf{Y}$, respectively.

The terms collected in A have in common that they contain two first differentials on the same side of $\mathbf{\Pi}$ in Eq. (43). Likewise, B is obtained by collecting terms that have two first differentials on different sides of $\mathbf{\Pi}$ in Eq. (43). Finally, C contains the terms with second differentials. The terms with $\mathrm{d}\mathbf{R}_\mathrm{u}$ are omitted at this early stage: they vanish after taking expectation because the entire correlation between $\tilde{\mathbf{Q}}$ and $\tilde{\mathbf{R}}_\mathrm{u}$ is accounted for by $\hat{\mathbf{H}}_\mathrm{u}$ and $\hat{\mathbf{Y}}_\mathrm{u}$.

Using straightforward manipulations, it can be shown that, owing to extensive cancellation, B is negligible with respect to A. (This is the reason why the terms containing $\mathbf{D}_\mathrm{H}$ and $\mathbf{D}_\mathrm{Y}$ are not further worked out.) Likewise, the four terms in C do not contribute, which is proved as follows. The estimated profiles must reproduce the SVD fit of $\tilde{\mathbf{Q}}$, see Eq. (32). It is explained in Section 2.1 in Part 2 that this fit is nearly unbiased if the signal-to-noise ratio is sufficiently high. Thus, any bias in the column profiles must be offset by the bias in the row profiles or, in other words, the profiles are interdependent. This suggests that decomposing the true data matrix $\mathbf{Q}$ as

$$
\begin{aligned}
\mathbf{Q} &= \left(\mathbf{H} + B(\hat{\mathbf{H}})\right)\left(\mathbf{Y} + B(\hat{\mathbf{Y}})\right)^\mathrm{T}\\
&\approx \mathbf{H}\mathbf{Y}^\mathrm{T} + \mathbf{H}\,B(\hat{\mathbf{Y}})^\mathrm{T} + B(\hat{\mathbf{H}})\mathbf{Y}^\mathrm{T}
\end{aligned}
\tag{45}
$$

could yield expressions for this bias. Replacing the bias in Eq. (45) by the second differential and rearranging yields

$$
\begin{aligned}
\mathrm{d}^2\mathbf{H} &\approx -\mathbf{H}(\mathrm{d}^2\mathbf{Y}^\mathrm{T})\mathbf{Y}^{+\mathrm{T}}\\
\mathrm{d}^2\mathbf{Y} &\approx -\mathbf{Y}(\mathrm{d}^2\mathbf{H}^\mathrm{T})\mathbf{H}^{+\mathrm{T}}
\end{aligned}
\tag{46}
$$

Unfortunately, the current approach does not lead to a closed form expression for bias in the profiles. It does, however, allow proving that C does not contribute. Inserting the expression for $\mathrm{d}^2\mathbf{Y}$ into the expression for $\mathrm{d}^2\mathbf{H}$ and vice versa gives

$$
\begin{aligned}
\mathrm{d}^2\mathbf{H} &\approx \mathbf{P}_\mathrm{H}\mathrm{d}^2\mathbf{H}\\
\mathrm{d}^2\mathbf{Y} &\approx \mathbf{P}_\mathrm{Y}\mathrm{d}^2\mathbf{Y}
\end{aligned}
\tag{47}
$$

which implies that $\mathrm{d}^2\mathbf{H}$ and $\mathrm{d}^2\mathbf{Y}$ lie in the column space of $\mathbf{H}$ and $\mathbf{Y}$, respectively, so that

$$
\begin{aligned}
\mathrm{d}^2\mathbf{H} &\approx \mathbf{H}\mathbf{T}_\mathrm{H}\\
\mathrm{d}^2\mathbf{Y} &\approx \mathbf{Y}\mathbf{T}_\mathrm{Y}
\end{aligned}
\tag{48}
$$

where $\mathbf{T}_\mathrm{H}$ and $\mathbf{T}_\mathrm{Y}$ are non-singular. Combining Eqs. (47) and (48) yields that

$$
\mathrm{d}^2\mathbf{H} \approx -\mathbf{H}\mathbf{T}_\mathrm{Y}^\mathrm{T}
\tag{49}
$$

and comparison of Eqs. (48) and (49) shows that

$$
\mathbf{T}_\mathrm{H} = -\mathbf{T}_\mathrm{Y}^\mathrm{T}
\tag{50}
$$

Using Eqs. (48) and (50), the first two terms of C can be expressed as

$$\begin{aligned} -\mathbf{H}^{+}(\mathrm{d}^2\mathbf{H})\mathbf{\Pi} &\approx -\mathbf{T}_{\mathrm{H}}\mathbf{\Pi} = \mathbf{T}_{\mathrm{Y}}^{\mathrm{T}}\mathbf{\Pi} \\ -\mathbf{\Pi}(\mathrm{d}^2\mathbf{Y}^{\mathrm{T}})\mathbf{Y}^{+\mathrm{T}} &\approx -\mathbf{\Pi}\mathbf{T}_{\mathrm{Y}}^{\mathrm{T}} \end{aligned} \tag{51}$$

and, since $\mathrm{diag}(\mathbf{T}_{\mathrm{Y}}^{\mathrm{T}}\mathbf{\Pi}) = \mathrm{diag}(\mathbf{\Pi}\mathbf{T}_{\mathrm{Y}}^{\mathrm{T}})$, these terms do not contribute to the diagonal of $\mathrm{d}^2\mathbf{\Pi}$. The same reasoning applies to the last two terms of C, owing to the special construction of $\hat{\mathbf{H}}_{\mathrm{u}}$ and $\hat{\mathbf{Y}}_{\mathrm{u}}$.

Thus, within the currently used approximation, one only has to deal with the first 12 terms after taking the expectation of Eq. (44). To work out terms 3 through 6 and 9 through 12, we need to bring the constant matrices outside the expectation operator. This can be done using [22]

$$E\big((\mathrm{d}\mathbf{H}^{\mathrm{T}})\mathbf{P}_{\mathrm{H}}\mathrm{d}\mathbf{H}\big) = KJ_1^{-1}E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H}) \tag{52}$$

and

$$E\big((\mathrm{d}\mathbf{H})\mathbf{H}^{+}\mathrm{d}\mathbf{H}\big) = J_1^{-1}\mathbf{H}^{+\mathrm{T}}E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H}) \tag{53}$$

where d$\mathbf{H}$ stands for d$\mathbf{H}$ or d$\mathbf{H}_{\mathrm{u}}$. Likewise,

$$E\big((\mathrm{d}\mathbf{Y}^{\mathrm{T}})\mathbf{P}_{\mathrm{Y}}\mathrm{d}\mathbf{Y}\big) = KJ_2^{-1}E(\mathrm{d}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{Y}) \tag{54}$$

and

$$\begin{aligned} E\big((\mathrm{d}\mathbf{Y}^{\mathrm{T}})\mathbf{Y}^{+\mathrm{T}}\mathrm{d}\mathbf{Y}^{\mathrm{T}}\big) &= E\big((\mathrm{d}\mathbf{Y})\mathbf{Y}^{+}\mathrm{d}\mathbf{Y}\big)^{\mathrm{T}} \\ &= J_2^{-1}\big(\mathbf{Y}^{+\mathrm{T}}E(\mathrm{d}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{Y})\big)^{\mathrm{T}} \\ &= J_2^{-1}E(\mathrm{d}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{Y})\mathbf{Y}^{+} \end{aligned} \tag{55}$$

where d$\mathbf{Y}$ stands for d$\mathbf{Y}$ or d$\mathbf{Y}_{\mathrm{u}}$. The validity of Eqs. (52)–(55) is easily verified by equating a particular matrix element, and using that $\mathrm{Tr}(\mathbf{P}_{\mathrm{H}}) = \mathrm{Tr}(\mathbf{P}_{\mathrm{Y}}) = K$, the dimension of the column space of $\mathbf{H}$ and $\mathbf{Y}$.

The result is that taking expectation of Eq. (44) yields that

$$\begin{aligned} B(\hat{\mathbf{\Pi}}) \approx\ & J_1^{-1}(J_1 - K - 1)\big[\mathbf{B}_{\mathrm{H}}^{-1}E\big(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H}_{\mathrm{u}}\big)\mathbf{\Pi} \\ & - \mathbf{B}_{\mathrm{H}}^{-1}E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H})\mathbf{\Pi}\big] \\ & + J_2^{-1}(J_2 - K - 1)\big[\mathbf{\Pi}E\big(\mathrm{d}\mathbf{Y}_{\mathrm{u}}^{\mathrm{T}}\mathrm{d}\mathbf{Y}\big)\mathbf{B}_{\mathrm{Y}}^{-1} \\ & - \mathbf{\Pi}E(\mathrm{d}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{Y})\mathbf{B}_{\mathrm{Y}}^{-1}\big] \end{aligned} \tag{56}$$

The next step consists of working out the expectation of the cross products of the first differentials. At this point, two—quite different—approximations have been considered. The first one, which seems to be the natural choice, is to insert the 'full linearisation' results derived in Section 2.3 of Part 2. The second one, which has led to Eq. (10), amounts to keeping only the first term in Eq. (13) of Part 2. The justification for doing so is that the remaining terms tend to cancel. It can be shown that the second approximation is equivalent to neglecting the uncertainty in $\hat{\mathbf{Y}}^{+\mathrm{T}}$ and $\hat{\mathbf{H}}^{+\mathrm{T}}$ in Eq. (33), which yields

$$\begin{aligned} \mathrm{d}\mathbf{H} &\approx (\mathrm{d}\mathbf{Q})\mathbf{Y}^{+\mathrm{T}} \\ \mathrm{d}\mathbf{Y} &\approx (\mathrm{d}\mathbf{Q}^{\mathrm{T}})\mathbf{H}^{+\mathrm{T}} \end{aligned} \tag{57}$$

From the special construction of $\hat{\mathbf{H}}_{\mathrm{u}}$ and $\hat{\mathbf{Y}}_{\mathrm{u}}$, it follows that

$$\begin{aligned} \mathrm{d}\mathbf{H}_{\mathrm{u}} &\approx (\mathrm{d}\mathbf{R}_{\mathrm{u}})\mathbf{Y}^{+\mathrm{T}}\mathbf{\Pi}^{-1} \\ \mathrm{d}\mathbf{Y}_{\mathrm{u}} &\approx \big(\mathrm{d}\mathbf{R}_{\mathrm{u}}^{\mathrm{T}}\big)\mathbf{H}^{+\mathrm{T}}\mathbf{\Pi}^{-1} \end{aligned} \tag{58}$$

Eq. (57) only differs from Eq. (58) with respect to the data matrix and the scaling by the inverse eigenvalue matrix. It is important to note that a division by zero will not occur because the inverse eigenvalue matrix cancels in Eq. (56). The expectation of the cross product of the first differential for $\mathbf{H}$ gives

$$\begin{aligned} E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H}) &\approx \mathbf{Y}^{+}E(\mathrm{d}\mathbf{Q}^{\mathrm{T}}\mathrm{d}\mathbf{Q})\mathbf{Y}^{+\mathrm{T}} \\ &= J_1V(\tilde{\mathbf{Q}})\mathbf{B}_{\mathrm{Y}}^{-1} \end{aligned} \tag{59}$$

since $E(\mathrm{d}\mathbf{Q}^{\mathrm{T}}\mathrm{d}\mathbf{Q}) = J_1V(\tilde{\mathbf{Q}})\mathbf{I}_{J_2}$ and $\mathbf{Y}^{+}\mathbf{Y}^{+\mathrm{T}} = \mathbf{B}_{\mathrm{Y}}^{-1}$. Likewise,

$$\begin{aligned} E(\mathrm{d}\mathbf{Y}^{\mathrm{T}}\mathrm{d}\mathbf{Y}) &\approx \mathbf{H}^{+}E(\mathrm{d}\mathbf{Q}\mathrm{d}\mathbf{Q}^{\mathrm{T}})\mathbf{H}^{+\mathrm{T}} \\ &= J_2V(\tilde{\mathbf{Q}})\mathbf{B}_{\mathrm{H}}^{-1} \end{aligned} \tag{60}$$

$$\begin{aligned} E\big(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H}_{\mathrm{u}}\big) &\approx \mathbf{Y}^{+}E\big(\mathrm{d}\mathbf{Q}^{\mathrm{T}}\mathrm{d}\mathbf{R}_{\mathrm{u}}\big)\mathbf{Y}^{+\mathrm{T}}\mathbf{\Pi}^{-1} \\ &= J_1V\big(\tilde{\mathbf{R}}_{\mathrm{u}}\big)\mathbf{B}_{\mathrm{Y}}^{-1}\mathbf{\Pi}^{-1} \end{aligned} \tag{61}$$

$$\begin{aligned} E\big(\mathrm{d}\mathbf{Y}_{\mathrm{u}}^{\mathrm{T}}\mathrm{d}\mathbf{Y}\big) &\approx \mathbf{\Pi}^{-1}\mathbf{H}^{+}E\big(\mathrm{d}\mathbf{R}_{\mathrm{u}}\mathrm{d}\mathbf{Q}^{\mathrm{T}}\big)\mathbf{H}^{+\mathrm{T}} \\ &= J_2V\big(\tilde{\mathbf{R}}_{\mathrm{u}}\big)\mathbf{\Pi}^{-1}\mathbf{B}_{\mathrm{H}}^{-1} \end{aligned} \tag{62}$$

Inserting Eqs. (59)–(62) in Eq. (56) yields that

$$B(\hat{\mathbf{\Pi}}) \approx J_1^{-1}(J_1 - K - 1)(J_1 - K) \times \left[V(\tilde{\mathbf{R}}_u)\mathbf{\Psi} - V(\tilde{\mathbf{Q}})\mathbf{\Psi}\mathbf{\Pi}\right] + J_2^{-1}(J_2 - K - 1)(J_2 - K) \times \left[V(\tilde{\mathbf{R}}_u)\mathbf{\Psi} - V(\tilde{\mathbf{Q}})\mathbf{\Pi}\mathbf{\Psi}\right] \quad (63)$$

where $\mathbf{\Psi} = (\mathbf{H}^T\mathbf{H})^{-1}(\mathbf{Y}^T\mathbf{Y})^{-1}$. Since $\mathrm{diag}(\mathbf{\Psi}\Pi) = \mathrm{diag}(\mathbf{\Pi}\mathbf{\Psi})$, Eq. (63) simplifies when focussing on a specific eigenvalue. The final simplifications that lead, after rearrangement, to Eq. (10), are obtained by inserting $V(\tilde{\mathbf{Q}}) = (1 + \alpha^2)V(\tilde{\mathbf{R}}_u)$ and $V(\tilde{\mathbf{R}}_u) = V(\tilde{\mathbf{R}})$.

It is reiterated that Eq. (63) is obtained by inserting a crude approximation in Eq. (56), rather than the full linearisation results. Interestingly, Denham [23] has reported a similar simplification to work well when applied to the covariance matrix of the regression coefficient estimates obtained by partial least squares (PLS).

Appendix B. Variance in bias obtained using Eqs. (17) and (18)

Only the final expressions are given here, since they are easily derived using earlier obtained results.

The variance in the bias in the eigenvalue, calculated using Eq. (17), is approximated by

$$V\left(\hat{B}(\hat{\pi})\right) \approx \left(\frac{\partial B(\hat{\pi})}{\partial \alpha}\right)^2 V(\hat{\alpha}) + \left(\frac{\partial B(\hat{\pi})}{\partial \psi}\right)^2 V(\hat{\psi}) + \left(\frac{\partial B(\hat{\pi})}{\partial \pi}\right)^2 V(\hat{\pi}) + \left(\frac{\partial B(\hat{\pi})}{\partial V(\tilde{\mathbf{R}})}\right)^2 V\left(\hat{V}(\tilde{\mathbf{R}})\right) \quad (64)$$

There is no term that accounts for the uncertainty in $\hat{\omega}$. This is justified by the consideration that estimating the correct model dimensionality is not a difficult task (see the discussion following Eq. (16)). In addition, often for data taken on modern instruments, $J_1 \gg F$ as well as $J_2 \gg F$, so that the actual value of F has little effect on $\hat{\omega}$.

The first term does not contribute (see the discussion following Eq. (22)), while the remaining partial derivatives are found as

$$\frac{\partial B(\hat{\pi})}{\partial \psi} = g^{-1}\omega\left[1 - (1 + \alpha^2)(\pi - B(\hat{\pi}))\right]V(\tilde{\mathbf{R}})$$

$$\frac{\partial B(\hat{\pi})}{\partial \pi} = -g^{-1}\omega\psi(1 + \alpha^2)V(\tilde{\mathbf{R}})$$

$$\frac{\partial B(\hat{\pi})}{\partial V(\tilde{\mathbf{R}})} = g^{-1}\omega\psi\left[1 - (1 + \alpha^2)(\pi - B(\hat{\pi}))\right]$$

where $g = 1 - \omega\psi(1 + \alpha^2)V(\tilde{\mathbf{R}})$.

In addition, three variances are required. Tedious manipulations show that

$$V(\hat{\psi}) \approx V(\tilde{\mathbf{Q}})\mathbf{J}\mathbf{J}^T \quad (65)$$

where $\mathbf{J} = -2(\mathbf{j}^T\mathbf{\Psi}\mathbf{Y}^T \otimes \mathbf{j}^T\mathbf{\Psi}\mathbf{H}^T) - (\mathbf{j}^T\mathbf{Y}^+ \otimes \mathbf{j}^T\mathbf{\Psi}\mathbf{H}^+) - (\mathbf{j}^T\mathbf{\Psi}\mathbf{Y}^+ \otimes \mathbf{j}^T\mathbf{H}^+)$ with $\mathbf{j}$ the $K \times 1$ vector with a one on the position associated with the analyte of interest and zeros otherwise. The variance in the eigenvalue is given by Eq. (14). Finally, assuming that $V(\tilde{\mathbf{R}})$ is distributed proportional to a χ^2 random variable with ν degrees of freedom, its variance is found as

$$V\left(\hat{V}(\tilde{\mathbf{R}})\right) = (2/\nu)V(\tilde{\mathbf{R}})^2 \quad (66)$$

Likewise, the variance in the prediction bias, calculated using Eq. (18), is approximated by

$$V\left(\hat{B}(\hat{c}_u)\right) \approx \left(\frac{\partial B(\hat{c}_u)}{\partial c_0}\right)^2 V(\tilde{c}_0) + \left(\frac{\partial B(\hat{c}_u)}{\partial \pi}\right)^2 V(\hat{\pi}) + \left(\frac{\partial B(\hat{c}_u)}{\partial B(\hat{\pi})}\right)^2 V\left(\hat{B}(\hat{\pi})\right) \quad (67)$$

where the partial derivatives are found as

$$\frac{\partial B(\hat{c}_u)}{\partial c_0} = g^{-1}\alpha B(\hat{\pi})$$

$$\frac{\partial B(\hat{c}_u)}{\partial \pi} = g^{-1}(2 - 2\pi + B(\hat{\pi}))\, B(\hat{c}_u)$$

$$\frac{\partial B(\hat{c}_u)}{\partial B(\hat{\pi})} = g^{-1}(\alpha c_0 + (\pi - 1)\, B(\hat{c}_u))$$

in which $g = (1 - \pi)(1 - \pi + B(\hat{\pi}))$. For the variances, see Eq. (22).

Appendix C. Generalisation of Eq. (10) to heteroscedastic and correlated measurement errors

Eq. (10) is derived under the iid assumption, which is usually too restrictive in applied work. Nevertheless, this equation is believed to be of considerable practical interest, since its generalisation to realistic error models is easily obtained without additional approximations. Only the main steps are outlined here to support the utility of Eq. (10). The practical implementation of the final result, which would by necessity be limited to a particular error model, is an interesting subject for future research.

First, it is observed that the assumption about the measurement errors enters the final stage of the derivation of Eq. (10), i.e., when inserting the results of Eqs. (59)–(62) in Eq. (56). The generalisation for Eq. (59) follows; the manipulations are similar for Eqs. (60)–(62). Starting from Eq. (57), and using the identity $\mathrm{vec}\mathbf{ABC} = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\mathrm{vec}\mathbf{B}$ [11], the first differential of the kth column of $\mathbf{H}$ is approximated as

$$\mathrm{d}\mathbf{h}_k \approx \mathbf{I}_{J_1}(\mathrm{d}\mathbf{Q})\mathbf{y}_k^{+} = \left(\mathbf{y}_k^{+\mathrm{T}} \otimes \mathbf{I}_{J_1}\right)\mathrm{vec}\,\mathrm{d}\mathbf{Q} \qquad (68)$$

where $\mathbf{y}_k^{+}$ is the kth column of $\mathbf{Y}^{+\mathrm{T}}$. Next, using straightforward manipulations involving the trace operator and the Kronecker product [11], a typical element of $E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H})$ is worked out as

$$\begin{aligned} E(\mathrm{d}\mathbf{H}^{\mathrm{T}}\mathrm{d}\mathbf{H})_{kl} &= E\left(\mathrm{d}\mathbf{h}_k^{\mathrm{T}}\mathrm{d}\mathbf{h}_l\right) \\ &\approx E\Big((\mathrm{vec}\,\mathrm{d}\mathbf{Q})^{\mathrm{T}}\left(\mathbf{y}_k^{+} \otimes \mathbf{I}_{J_1}\right) \\ &\quad \times \left(\mathbf{y}_l^{+\mathrm{T}} \otimes \mathbf{I}_{J_1}\right)\mathrm{vec}\,\mathrm{d}\mathbf{Q}\Big) \\ &= \mathrm{Tr}\left(\mathbf{V}(\tilde{\mathbf{Q}})\left(\mathbf{y}_k^{+}\mathbf{y}_l^{+\mathrm{T}} \otimes \mathbf{I}_{J_1}\right)\right) \qquad (69) \end{aligned}$$

where $\mathbf{V}(\tilde{\mathbf{Q}}) = E(\mathrm{vec}\,\mathrm{d}\mathbf{Q}(\mathrm{vec}\,\mathrm{d}\mathbf{Q})^{\mathrm{T}})$ denotes the full $J_1J_2 \times J_1J_2$ covariance matrix for the measurement errors in $\tilde{\mathbf{Q}}$.

Inserting the generalisations of Eqs. (59)–(62) in Eq. (56) will yield the desired generalisation of Eq. (10), without additional approximations. It is expected that the interpretability will be limited, unless a special structure of the errors allows for further simplification. Interpretability will be lost altogether if, instead of the crude approximation Eq. (57), the full linearisation result is used for $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$ (see Eqs. (11) and (13) in Part 2).

A final remark about the handling of heteroscedastic and correlated errors is in order. It is observed that the GRAM eigenvalue problem results after decomposing $\tilde{\mathbf{Q}}$ using the SVD. The SVD is a least-squares method, which implies that its optimal fitting properties are only retained in the presence of heteroscedastic and correlated measurement errors if the data are properly pre-treated. Following Paatero and Tapper [24], Faber et al. [2] have suggested to pre- and post-multiply the data with diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_r$, respectively. (For more details, see Refs. [2,24].) Using such a scaling technique, model Eq. (5) becomes

$$\tilde{\mathbf{Q}}^{\mathrm{scl}} = \mathbf{D}_1\tilde{\mathbf{Q}}\mathbf{D}_r = \mathbf{H}^{\mathrm{scl}}(\mathbf{Y}^{\mathrm{scl}})^{\mathrm{T}} + \mathbf{D}_1\mathbf{E}_Q\mathbf{D}_r$$

$$\tilde{\mathbf{R}}_u^{\mathrm{scl}} = \mathbf{D}_1\tilde{\mathbf{R}}_u\mathbf{D}_r = \mathbf{H}^{\mathrm{scl}}\mathbf{\Pi}(\mathbf{Y}^{\mathrm{scl}})^{\mathrm{T}} + \mathbf{D}_1\mathbf{E}_u\mathbf{D}_r \qquad (70)$$

where $\mathbf{H}^{\mathrm{scl}} = \mathbf{D}_1\mathbf{H}$ and $\mathbf{Y}^{\mathrm{scl}} = \mathbf{D}_r\mathbf{Y}$. Obviously, the scaling must be undone after solving the eigenvalue problem to obtain the estimates for $\mathbf{H}$ and $\mathbf{Y}$. In addition, the scaled measurement errors should enter the bias expressions derived in this paper. Data pre-treatment in the context of GRAM has received little attention in the literature, which leaves room for future research. Currently, the most detailed discussion of data-pretreatment related to SVD is given by Wentzell et al. [25].

Appendix D. The noise addition method for determining bias and variance

We have not been able to derive a closed-form expression for the bias in $\hat{\mathbf{H}}$ and $\hat{\mathbf{Y}}$, which is unsatisfactory. To determine (and correct for) the bias in the profile estimates, we have applied a simulation tech-

nique that is based on *adding noise* to the experimental data. The key assumption underlying this technique is that the response of the system of equations to noise remains (approximately) constant when adding noise. Clearly, the key assumption implies that the simulation technique can be used to determine bias as well as variance. Here, it will be used to demonstrate the validity of Eq. (51), while it is applied in Part 2 to test expressions for the variance in the predictions and profiles estimates. For an excellent discussion of the noise addition method applied to variance estimation, see Ref. [26]. Derks et al. [27] have reported successful variance estimation in the context of artificial neural networks. When applied to bias estimation, the method is known as simulation extrapolation (SIMEX) [28]. The principle is illustrated for the eigenvalue $\hat{\pi}$, which is a scalar. However, the generalisation to vectors and matrices is straightforward. Variance estimation is treated first, not only because it is simplest, but also because the result is used to determine the uncertainty of the bias estimate.

When applied to variance estimation, it is assumed that the variance in $\hat{\pi}$ can be approximated as

$$V(\hat{\pi}) \approx \upsilon V(\tilde{\mathbf{R}}) \tag{71}$$

which seems to be reasonable, because the squared expectation of the first-order term of a Taylor series expansion is the main contribution to variance. Likewise, adding noise to $\tilde{\mathbf{R}}_0$ and $\tilde{\mathbf{R}}_u$ with variance V_{add} will yield eigenvalues $\hat{\pi}_{add}$, with variance estimate

$$\hat{V}(\hat{\pi}_{add}) \approx \upsilon V_{add} \tag{72}$$

Eq. (72) will be a good approximation if the added noise is small enough. It is important to note that the noise already present in the data matrices does not contribute to the variance of $\hat{\pi}_{add}$, because it is constant during the noise additions (it is 'baked in'). Combining Eqs. (71) and (72) suggests that the variance of $\hat{\pi}$ can be estimated as

$$\hat{V}(\hat{\pi}) = \hat{V}(\hat{\pi}_{add}) \frac{\hat{V}(\tilde{\mathbf{R}})}{V_{add}} \tag{73}$$

For the current application, it is assumed that bias scales up linearly with variance, which is a simplification with respect to the example treated in Ref. [28]. This assumption seems to be reasonable for $\hat{\pi}$, but it must be verified for the profile estimates. Thus, the bias in $\hat{\pi}$ is approximated as

$$B(\hat{\pi}) \approx bV(\tilde{\mathbf{R}}) \tag{74}$$

Bias is different from variance in the sense that adding noise will yield a composite bias,

$$B(\hat{\pi}_{add}) \approx B(\hat{\pi}) + bV_{add} \tag{75}$$

It is observed that, unlike Eq. (71), Eq. (75) is an equation with two unknowns, $B(\hat{\pi})$ and b, which implies that noise should be added at more than one level. Adding noise at two levels,[3] V_1 and V_2, gives average eigenvalues, $\hat{E}(\hat{\pi}_1)$ and $\hat{E}(\hat{\pi}_2)$, which can be approximated as

$$\begin{aligned} \hat{E}(\hat{\pi}_1) &\approx \pi + B(\hat{\pi}) + bV_1 \\ \hat{E}(\hat{\pi}_2) &\approx \pi + B(\hat{\pi}) + bV_2 \end{aligned} \tag{76}$$

Combining Eqs. (74) and (76) suggests that bias can be estimated as

$$\hat{B}(\hat{\pi}) = \frac{\hat{E}(\hat{\pi}_2) - \hat{E}(\hat{\pi}_1)}{V_2 - V_1} \hat{V}(\tilde{\mathbf{R}}) \tag{77}$$

Eq. (77) has been successfully used, in a form adapted to vectors, to correct the profile estimates for bias (see Fig. 4).

It is important to note that Eq. (77) is an alternative to Eq. (17). The foremost criteria for assessing the utility of bias estimates are their interpretability and relative uncertainty (i.e., efficiency). With respect to interpretability, expressions certainly have an advantage. Knowing which factors influence bias enables one to set up the experiments in such a way that a significant bias can be avoided altogether. With respect to the uncertainty of the bias estimate, simulation techniques could have an advantage. The uncer-

[3] In general, the relationship between noise and bias will be more complicated so that one needs to add noise at more than two levels. For more details, see Ref. [28].

tainty in the estimate obtained from Eq. (77) is obtained from a first-order approximation as

$$
\begin{aligned}
&V\left(\hat{B}(\hat{\pi})\right) \\
&\quad \approx \left(\frac{V(\tilde{\mathbf{R}})}{V_2 - V_1}\right)^2 \left(V\left(\hat{E}(\hat{\pi}_2)\right) + V\left(\hat{E}(\hat{\pi}_1)\right)\right) \\
&\qquad + \left(\frac{E(\hat{\pi}_2) - E(\hat{\pi}_1)}{V_2 - V_1}\right)^2 V\left(\hat{V}(\tilde{\mathbf{R}})\right) \\
&\quad = \left(\frac{V(\tilde{\mathbf{R}})}{V_2 - V_1}\right)^2 \frac{1}{N_{\text{add}}}\left(V(\hat{\pi}_2) + V(\hat{\pi}_1)\right) \\
&\qquad + \left(\frac{E(\hat{\pi}_2) - E(\hat{\pi}_1)}{V_2 - V_1}\right)^2 \frac{2}{V} V(\tilde{\mathbf{R}})^2 \\
&\quad \approx \frac{(V_2 + V_1)V(\tilde{\mathbf{R}})}{(V_2 - V_1)^2 N_{\text{add}}} V(\hat{\pi}) \\
&\qquad + \left(\frac{E(\hat{\pi}_2) - E(\hat{\pi}_1)}{V_2 - V_1}\right)^2 \frac{2}{V} V(\tilde{\mathbf{R}})^2 \qquad (78)
\end{aligned}
$$

where, in the second step, the variance in the averages is replaced by the variance in a single replication, divided by the number of noise additions (N_{add}), and the χ^2 assumption is applied to $\hat{V}(\tilde{\mathbf{R}})$ (see Eq. (66)). The last step results from inserting Eq. (73).

Eq. (78) is useful for determining reasonable values for N_{add}, V_1 and V_2. A sound criterion is that $V(\hat{B}(\hat{\pi}))$ should be smaller than $V(\hat{\pi})$ by a factor 10 (say); otherwise, the bias correction introduces an unduly large uncertainty. Ignoring the term associated with the uncertainty in $\hat{V}(\tilde{\mathbf{R}})$, which is reasonable if V is large, yields the condition that N_{add} should exceed $10V(\tilde{\mathbf{R}})(V_2 + V_1)/(V_2 - V_1)^2$. Finally, choosing $V_2 = 2V_1$ yields the requirement $N_{\text{add}} > 20V(\tilde{\mathbf{R}})/V_1$. The noise additions discussed in this paper are carried out with $V(\tilde{\mathbf{R}})/V_1 = 5$, 10 and 20. It follows that $N_{\text{add}} = 1000$ is a reasonable choice for all noise levels, provided that adding noise at two levels is sufficient of course. It turns out that this is only the case for the lowest noise level.

References

[1] E. Sánchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496–499.
[2] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 95–109.
[3] K. Booksh, B.R. Kowalski, J. Chemom. 8 (1994) 45–63.
[4] N.M. Faber, L.M.C. Buydens, G. Kateman, J. Chemom. 8 (1994) 181–203.
[5] S. Van Huffel, J. Vandewalle, The Total Least Squares Problem. Computational Aspects and Analysis, SIAM, Philadelphia, 1991.
[6] J. Riu, F.X. Rius, Trends Anal. Chem. 16 (1997) 211–216.
[7] N.M. Faber, L.M.C. Buydens, G. Kateman, J. Chemom. 7 (1993) 495–526.
[8] K. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11 (1997) 419–461.
[9] A. Lorber, Anal. Chem. 58 (1986) 1167–1172.
[10] N.M. Faber, R. Boqué, J. Ferré, Chemom. Intell. Lab. Syst. 55 (2001) 91–100.
[11] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, Chichester, 1988.
[12] H.A.L. Kiers, A.K. Smilde, J. Chemom. 9 (1995) 179–195.
[13] S.E. Leurgans, R.T. Ross, R.B. Abel, SIAM J. Matrix Anal. Appl. 14 (1993) 1064–1083.
[14] N.M. Faber, L.M.C. Buydens, G. Kateman, J. Chemom. 8 (1994) 147–154.
[15] N.M. Faber, Anal. Lett. 32 (1999) 2899–2906.
[16] M. Linder, R. Sundberg, J. Chemom., in press.
[17] F.P. Zscheile, H.C. Murray, G.A. Baker, R.G. Peddicord, Anal. Chem. 34 (1962) 1776–1780.
[18] C. Ritter, J.A. Gilliard, J. Cumps, B. Tilquin, Anal. Chim. Acta 318 (1995) 125–136.
[19] J.-H. Wang, P.K. Hopke, Anal. Chim. Acta 412 (2000) 177–184.
[20] S. de Jong, H.A.L. Kiers, Chemom. Intell. Lab. Syst. 14 (1992) 155–164.
[21] M.G. Moran, B.R. Kowalski, Anal. Chem. 56 (1984) 562–569.
[22] S.D. Hodges, P.G. Moore, Appl. Stat. 21 (1972) 185–195.
[23] M.C. Denham, J. Chemom. 14 (2000) 351–361.
[24] P. Paatero, U. Tapper, Chemom. Intell. Lab. Syst. 18 (1993) 183–194.
[25] P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, B.R. Kowalski, J. Chemom. 11 (1997) 339–366.
[26] W.H. Press, B.P. Flannery, S.A. Teukolski, W.T. Vetterling, Numerical Recipes. The Art of Scientific Computing, Cambridge Univ. Press, Cambridge, 1988.
[27] E.P.P.A. Derks, M.S. Sánchez Pastor, L.M.C. Buydens, Chemom. Intell. Lab. Syst. 28 (1995) 49–60.
[28] R.J. Carroll, D. Ruppert, L.A. Stefanski, Measurement Error in Nonlinear Models, Chapman & Hall, London, 1995.