Chapter IV Multiresponse parameter estimation

IV.1 Introduction

Multiresponse parameter estimation is ubiquitous, however, it is not always recognized as such. Measurements which are a function of many independent variables lead to multivariate analysis, which is defined as “the body of statistical methodology used to analyse simultaneous measurements on many variables”. Most of these methods are based upon the multivariate normal distribution. With linear models it can be shown that the multivariate Gauss-Markoff model falls apart into univariate Gauss-Markoff models, and can be treated in that way. With nonlinear models a distinction has to be made. With independent observations the multiresponse model can again be treated as a simple nonlinear regression model. Otherwise, a more complicated multiresponse estimation technique becomes necessary. We will pay special attention to the treatment of conditionally linear parameters in nonlinear models.

IV.2 The multiresponse model

Until now we have considered one vector $\underline{y}$ consisting of n observations $\underline{y}_i$. However, in many experiments several response vectors are measured simultaneously. Consider the model

$$\underline{y}_{ij} = g_j(x_i|\theta) + \underline{\nu}_{ij} \qquad \text{Eq. IV.1}$$

where the index j denotes the j-th response vector, and g_j denotes the model function for the j-th response. Note that the observations $\underline{y}_{ij}$, $j = 1, \ldots, m$ share the independent variables x_i. When g_j is linear in the unknown parameters θ we have

$$\underline{y}_{ij} = \sum_{l=1}^{k} h_l(\xi_i)\theta_{lj} + \underline{\nu}_{ij} \quad \text{or} \qquad \text{Eq. IV.2}$$

$$\underline{y}_{.i} = X\theta_{.j} + \underline{\nu}_{.i} \quad \text{or} \quad \underline{Y} = X\Theta + \underline{\nu}_{..}$$

where the subscript $_{.j}$ denotes the column vector containing the n observations $\underline{y}_{ij}$, or the k parameters θ_{lj}. We make the following assumptions

(a) $\underline{\nu}_{.j} \sim N(0, \sigma_{jj}I)$, thus observations within a column vector $\underline{y}_{.j}$ are independent, and $D(\underline{y}_{.j}) = \sigma_{jj}I$. When this condition is not fulfilled, a transformation as described in Eq. II.3 has to be done first.

(b) The cross-covariance between $\underline{y}_{.j}$ and $\underline{y}_{.l}$ is given by

$$D(\underline{y}_{.j}, \underline{y}_{.l}) = \sigma_{jl} I_n \qquad D(\mathrm{vec}Y) = \Sigma \otimes I_n \qquad \text{Eq. IV.3}$$

where the $m \times m$ matrix $\Sigma = (\sigma_{jl})$. In the special case that the different responses are independent, $\sigma_{jl} = \sigma_{jj}\delta_{jl}$, the model is equivalent to a single response model, with all observations $\underline{y}_{.j}$ concatenated, and the methods of Chapter II and Chapter III can directly be applied. The right hand side of Eq. IV.3 is explained in the following

Intermezzo IV.1. Kronecker product and vector representation of a matrix

Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = (b_{ij})$ an $p \times q$ matrix, then the Kronecker product $A \otimes B$ is defined by the $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \dots & \dots & \dots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix} \qquad \text{Eq. IV.4}$$

Theorem IV.1. The following properties of the Kronecker product hold

$$(A \otimes B)^T = A^T \otimes B^T \qquad \text{Eq. IV.5}$$

$$(A + B) \otimes C = A \otimes C + B \otimes C \qquad \text{Eq. IV.6}$$

Let A,C,B,D denote, respectively, $l \times m$, $m \times n$, $p \times q$ and $q \times r$ matrices, then

$$(A \otimes B)(C \otimes D) = AC \otimes BD \qquad \text{Eq. IV.7}$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \qquad \text{Eq. IV.8}$$

Proof: Using the definition in Eq. IV.4 we find

$$(A \otimes B)^T = \begin{bmatrix} a_{11}B^T & \dots & a_{m1}B^T \\ \dots & \dots & \dots \\ a_{1n}B^T & \dots & a_{mn}B^T \end{bmatrix} = A^T \otimes B^T .$$

$$(A + B) \otimes C = \begin{bmatrix} (a_{11} + b_{11})B & \dots & (a_{1n} + b_{1n})B \\ \dots & \dots & \dots \\ (a_{m1} + b_{m1})B & \dots & (a_{mn} + b_{mn})B \end{bmatrix} = A \otimes C + B \otimes C .$$

$$(A \otimes B)(C \otimes D) = \begin{bmatrix} a_{11}B & \dots & a_{1m}B \\ \dots & \dots & \dots \\ a_{l1}B & \dots & a_{lm}B \end{bmatrix} \begin{bmatrix} c_{11}D & \dots & c_{1n}D \\ \dots & \dots & \dots \\ c_{m1}D & \dots & c_{mn}D \end{bmatrix} = \sum_{i=1}^{m} \begin{bmatrix} a_{1i}c_{i1}BD & \dots & a_{1i}c_{in}BD \\ \dots & \dots & \dots \\ a_{li}c_{i1}BD & \dots & a_{li}c_{in}BD \end{bmatrix} =$$

$AC \otimes BD$. Let $C = A^{-1} \otimes B^{-1}$ then with Eq. IV.7 $(A \otimes B)C = I$ thus Eq. IV.8 follows.□

The vector representation $\mathrm{vec}A$ originates by writing the columns of A one below the other:

$$\mathrm{vec}A = (a_{11}, \dots, a_{m1}, a_{12}, \dots, a_{mn})^T \qquad \text{Eq. IV.9}$$

Theorem IV.2. Let A,B,C denote, respectively, $m \times n$, $n \times p$ and $p \times s$ matrices. Then

$$\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}B \qquad \text{Eq. IV.10}$$

Let A and B be two $n \times n$ matrices. Then

$$(\mathrm{vec}A)^T\mathrm{vec}B = (\mathrm{vec}B)^T\mathrm{vec}A = \mathrm{trace}(AB^T) \qquad \text{Eq. IV.11}$$

Let A,B,C denote, respectively, $m \times n$, $n \times n$ and $m \times m$ matrices. Then

$$(\mathrm{vec}A)^T(B \otimes C)\mathrm{vec}A = \mathrm{trace}(ABA^TC^T) = \mathrm{trace}(AB^TA^TC) \qquad \text{Eq. IV.12}$$

Proof: $\mathrm{vec}(ABC) = \sum_{i=1}^{p} \begin{bmatrix} c_{i1}Ab_{.i} \\ \dots \\ c_{is}Ab_{.i} \end{bmatrix} = \begin{bmatrix} c_{11}A & \dots & c_{p1}A \\ \dots & \dots & \dots \\ c_{1s}A & \dots & c_{ps}A \end{bmatrix} \begin{bmatrix} b_{.1} \\ \dots \\ b_{.p} \end{bmatrix} = (C^T \otimes A)\mathrm{vec}B$.

$(\mathrm{vec}A)^T\mathrm{vec}B = \sum_{i,\,i} a_{ij}b_{ij} = \sum_{i,\,i} a_{ij}b_{ji}^T = \mathrm{trace}(AB^T)$. Application of Eq. IV.10 and Eq. IV.11 to Eq. IV.12 gives: $(\mathrm{vec}A)^T(B \otimes C)\mathrm{vec}A = (\mathrm{vec}A)^T\mathrm{vec}(CAB^T) =$

$\mathrm{trace}(ABA^TC^T) = (\mathrm{vec}(CAB^T))^T\mathrm{vec}A = \mathrm{trace}(CAB^TA^T) = \mathrm{trace}(AB^TA^TC)$.□

IV.3 Multivariate Gauss-Markoff model

Eq. IV.2 and Eq. IV.3 define the multivariate Gauss-Markoff model, which we solve with the aid of Intermezzo IV.1. This model falls apart into the following univariate Gauss-Markoff model

$$E[\mathrm{vec}\underline{Y}] = (I_m \otimes X)\mathrm{vec}\Theta \qquad D(\mathrm{vec}Y) = \Sigma \otimes I_n \qquad \text{Eq. IV.13}$$

with as solution (see Eq. II.9)

$$\mathrm{vec}\underline{\hat{\Theta}} = ((I_m \otimes X)^T(\Sigma \otimes I_n)^{-1}(I_m \otimes X))^- (I_m \otimes X)^T(\Sigma \otimes I_n)^{-1}\mathrm{vec}\underline{Y} \qquad \text{Eq. IV.14}$$

We apply Eq. IV.5 and Eq. IV.6 to find $(I_m \otimes X)^T(\Sigma \otimes I_n)^{-1}(I_m \otimes X) = \Sigma^{-1} \otimes X^T X$.

Finally, using Eq. IV.7 and Eq. IV.6: $(\Sigma^{-1} \otimes X^T X)^- (\Sigma^{-1} \otimes X^T)\mathrm{vec}\underline{Y} =$

$(\Sigma \otimes (X^T X)^-)(\Sigma^{-1} \otimes X^T)\mathrm{vec}\underline{Y} = (I_m \otimes (X^T X)^- X^T)\mathrm{vec}\underline{Y}$ which is equivalent to

$$\underline{\hat{\theta}}_{.j} = (X^T X)^- X^T \underline{y}_{.j} \qquad \text{or} \qquad \underline{\hat{\Theta}} = (X^T X)^- X^T \underline{Y} \qquad \text{Eq. IV.15}$$

possessing cross- and auto-covariances (using Eq. IV.3)

$$D(\underline{\hat{\theta}}_{.j}, \underline{\hat{\theta}}_{.l}) = \sigma_{jl}(X^T X)^- X^T X((X^T X)^-)^T \qquad \text{Eq. IV.16}$$

The sum of squares $S(\Theta)$ of the residuals now is given by (using Eq. IV.12)

$$\begin{aligned} S(\Theta) &= (\mathrm{vec}(\underline{Y} - X\Theta))^T(\Sigma \otimes I_n)^{-1}(\mathrm{vec}(\underline{Y} - X\Theta)) \\ &= \mathrm{trace}((\underline{Y} - X\Theta)\Sigma^{-1}(\underline{Y} - X\Theta)^T) = \sum_{i,l} (\underline{y}_{.j} - X\theta_{.j})^T \sigma_{jl}^{-1}(\underline{y}_{.l} - X\theta_{.l}) \end{aligned} \qquad \text{Eq. IV.17}$$

where $\Sigma^{-1} = (\sigma_{jl}^{-1})$. $S(\Theta)$ attains a minimum when $\Theta = \underline{\hat{\Theta}}$. The covariance matrix Σ is estimated by (see Eq. II.29 and Eq. II.30)

$$\underline{\hat{\Sigma}} = \frac{(\underline{Y} - X\underline{\hat{\Theta}})^T(\underline{Y} - X\underline{\hat{\Theta}})}{n-q} = \frac{\underline{Y}^T(I - X(X^T X)^- X^T)\underline{Y}}{n-q} \qquad \text{Eq. IV.18}$$

To proof that $\underline{\hat{\Sigma}}$ is unbiased we need the following

Theorem IV.3. Let $x \sim N(\mu_x, \Sigma_{xx})$, $y \sim N(\mu_y, \Sigma_{yy})$ and $D(\underline{x}, \underline{y}) = \Sigma_{xy}$ then

$$E[\underline{x}^T A \underline{y}] = \mathrm{trace}(A\Sigma_{yx}) + \mu_x^T A \mu_y \qquad \text{Eq. IV.19}$$

Proof: $E[\underline{x}^T A \underline{y}] = E\left[\sum_{i,j} \underline{x}_i a_{ij} \underline{y}_j\right] = E[\mathrm{trace} A\underline{y}\underline{x}^T] = \mathrm{trace}\{AE[\underline{y}\underline{x}^T]\} =$

$\mathrm{trace}\{A(\Sigma_{yx} + \mu_y \mu_x^T)\} = \mathrm{trace}(A\Sigma_{yx}) + \sum_{i,i} \mu_{xi} a_{ij} \mu_{yj} = \mathrm{trace}(A\Sigma_{yx}) + \mu_x^T A \mu_y$.□

Now, using Eq. II.18, Eq. II.29 and Eq. IV.3 we get $E[\underline{\hat{\sigma}}_{jl}] = \frac{E[\underline{y}_{.j}^T(I - X(X^T X)^- X^T)\underline{y}_{.l}]}{n-q}$

$$= \frac{\mathrm{trace}\{(I - X(X^T X)^- X^T)\sigma_{lj}\}}{n-q} + \frac{(X\theta_{.j})^T(I - X(X^T X)^- X^T)X\theta_{.l}}{n-q} = \sigma_{lj} = \sigma_{jl} \quad .\square$$

*IV.4 Multiresponse nonlinear model**

To find the estimate for the nonlinear model we must determine the maximum likelihood criterion. Under the assumptions (a) and (b) the likelihood function is proportional to

$$\prod_{i=1}^{n} (\det\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(y_{i\cdot} - g_{\cdot}(x_i|\theta))\Sigma^{-1}(y_{i\cdot} - g_{\cdot}(x_i|\theta))^T\right) \qquad \text{Eq. IV.20}$$

thus apart from a normalizing constant the loglikelihood function to be maximized is

$$L(\theta, \Sigma^{-1}) = k + \frac{n}{2}\log(\det\Sigma^{-1}) - \frac{1}{2}\text{trace}\{(\underline{Y} - G(\theta))\Sigma^{-1}(\underline{Y} - G(\theta))^T\} \qquad \text{Eq. IV.21}$$

We will first maximize this expression with respect to Σ^{-1} for which we need the following

Theorem IV.4. The derivative of $\log(\det A)$ with respect to the matrix A is given by

$$\frac{\partial}{\partial A}\log(\det A) = (A^{-1})^T \qquad \text{Eq. IV.22}$$

Proof: Let a_{ij}^* denote the cofactor of a_{ij}: $a_{ij}^* = (-1)^{i+j} m_{ij}$ where the minor m_{ij} is defined as the determinant of A with the i-th row and j-th column deleted. It is well known that $a_{ji}^{-1} = (\det A)^{-1} a_{ij}^*$ (Cramer's rule for the calculation of the inverse matrix). Now using $\det A = \sum_i a_{ij} a_{ij}^*$ we get $\frac{\partial}{\partial a_{kl}}\log(\det A) = (\det A)^{-1}\frac{\partial}{\partial a_{kl}}\left(\sum_j a_{ij} a_{ij}^*\right) = \frac{a_{kl}^*}{\det A} = a_{lk}^{-1}$.□

Using Theorem IV.4 we find $\frac{\partial}{\partial(\Sigma^{-1})_{mn}}\log(\det\Sigma^{-1}) = \Sigma_{nm}$. Furthermore

$$\frac{\partial}{\partial\sigma_{mn}^{-1}}\text{trace}\{(\underline{Y} - G)\Sigma^{-1}(\underline{Y} - G)^T\} = \frac{\partial}{\partial\sigma_{mn}^{-1}}\text{trace}\{\Sigma^{-1}(\underline{Y} - G)^T(\underline{Y} - G)\}$$

$\frac{\partial}{\partial\sigma_{mn}^{-1}}\left\{\sum_{j,l}\sigma_{jl}^{-1}\{(\underline{Y} - G)^T(\underline{Y} - G)\}_{lj}\right\} = \{(\underline{Y} - G)^T(\underline{Y} - G)\}_{nm}$. Thus

$$\frac{\partial}{\partial\sigma_{mn}^{-1}}L(\theta, \Sigma^{-1}) = -\frac{n}{2}\sigma_{nm} + \frac{1}{2}\{(\underline{Y} - G)^T(\underline{Y} - G)\}_{nm} = 0 \qquad \text{Eq. IV.23}$$

The biased maximum likelihood estimate of Σ is given by

$$\hat{\underline{\Sigma}}_{ML} = \frac{(\underline{Y} - G)^T(\underline{Y} - G)}{n} \qquad \text{Eq. IV.24}$$

Substitution into Eq. IV.21 gives

$$L(\theta, \hat{\underline{\Sigma}}_{ML}) = k' - \frac{n}{2}\log(\det\hat{\underline{\Sigma}}_{ML}) = k'' - \frac{n}{2}\log(\det\{(\underline{Y}-G)^T(\underline{Y}-G)\}) \qquad \text{Eq. IV.25}$$

Thus maximum likelihood leads to the determinant criterion: minimize as function of θ the $\det\{(\underline{Y}-G)^T(\underline{Y}-G)\} \equiv \Upsilon$. It is straightforward to show that the univariate case leads to the least squares criterion.

*IV.4.1 Confidence intervals and confidence regions**

To draw inferences about parameters in the multiresponse nonlinear model we make a Taylor expansion of Υ around $\hat{\underline{\theta}}$

$$\Upsilon(\theta) = \Upsilon(\hat{\underline{\theta}}) + \frac{1}{2}(\theta-\hat{\underline{\theta}})^T\left(\frac{\partial^2}{\partial\theta^2}\Upsilon(\hat{\underline{\theta}})\right)(\theta-\hat{\underline{\theta}}) \qquad \text{Eq. IV.26}$$

Comparing Eq. IV.26 to Eq. II.69 of the linear model we see that the role of X^TX now has to be replaced by $\frac{1}{2}\frac{\partial^2}{\partial\theta^2}\Upsilon(\hat{\underline{\theta}})$. Analogously to Eq. II.64 and Eq. III.29 we define the approximate $100(1-\alpha)\%$ confidence limits for θ_j:

$$\hat{\underline{\theta}}_j \pm t_{n-q,\,\alpha/2}\hat{\underline{\sigma}}\sqrt{\left(\left(\frac{1}{2}\frac{\partial^2}{\partial\theta^2}\Upsilon(\hat{\underline{\theta}})\right)^{-1}\right)_{jj}} \qquad \text{Eq. IV.27}$$

with

$$\hat{\underline{\sigma}}^2 = \frac{\Upsilon(\hat{\underline{\theta}})}{n-q} \qquad \text{Eq. IV.28}$$

An approximate $100(1-\alpha)\%$ confidence region for the parameters is given by (compare Eq. II.71)

$$(\theta-\hat{\underline{\theta}})^T\left(\frac{1}{2}\frac{\partial^2}{\partial\theta^2}\Upsilon(\hat{\underline{\theta}})\right)(\theta-\hat{\underline{\theta}}) \le q\hat{\underline{\sigma}}^2F_{q,\,n-q,\,\alpha} \qquad \text{Eq. IV.29}$$

Note that in multiresponse models the number of degrees of freedom is $n-q$, the number of observations minus the number of parameters (see also Eq. IV.18). Furthermore n also has to be larger than m, the number of responses, since otherwise the determinant equals zero. This follows because $\text{rank}(\underline{Y}-G) \le \min(m, n)$ and thus $\text{rank}(\underline{Y}-G)^T(\underline{Y}-G) \le \min(m, n)$. When $n < m$ then $(\underline{Y}-G)^T(\underline{Y}-G)$ is rank deficient and $\Upsilon = 0$.

According to Bates and Watts (1988) near the optimum Υ is controlled by the linear combination of responses corresponding to the smallest singular value of $\underline{Y} - G$ and this vector has dimension n with $n - q$ degrees of freedom.

*IV.4.2 Case study: complex dielectric constant**

The model function for the complex dielectric constant ε^* reads

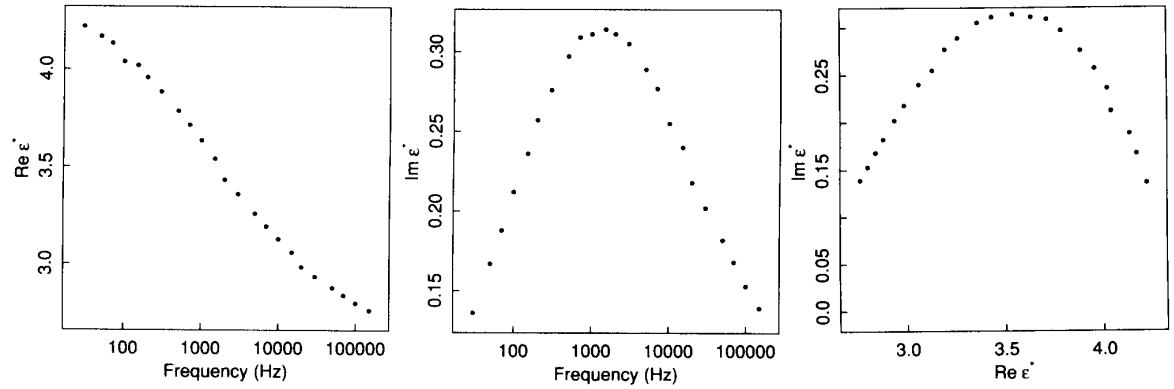

Fig.IV.1. Plots of the real and imaginary components of a polymer dielectric constant ε^* versus frequency and in the complex plane.

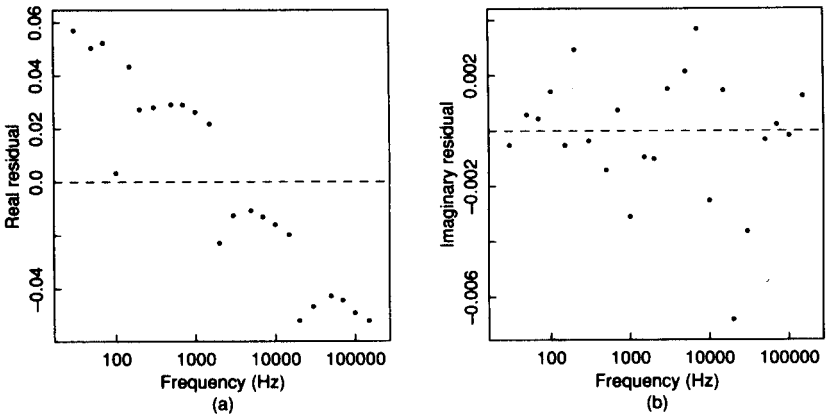

Fig.IV.2. Residuals from the initial fit.

$$\varepsilon^* = \varepsilon' - i\varepsilon'' = \varepsilon_\infty + \frac{\varepsilon_0 - \varepsilon_\infty}{\{1 + (2\pi i f/f_0)^\alpha\}^\beta} \qquad \text{Eq. IV.30}$$

and measurements of ε' and ε'' of a polymer are depicted in Fig.IV.1. which is taken from Bates and Watts (1988). Eq. IV.30 can be rewritten, abbreviating $\gamma = (2\pi f/f_0)^\alpha$, into

$$\begin{aligned} \varepsilon' &= \varepsilon_\infty + (\varepsilon_0 - \varepsilon_\infty)(1 + \gamma^2 + 2\gamma\cos(\alpha\pi/2))^{-\beta}\cos(\beta\phi) \\ \varepsilon'' &= (\varepsilon_0 - \varepsilon_\infty)(1 + \gamma^2 + 2\gamma\cos(\alpha\pi/2))^{-\beta}\sin(\beta\phi) \end{aligned} \qquad \text{Eq. IV.31}$$

with $\phi = \text{atan}\{(\gamma\sin(\alpha\pi/2))/(1+\gamma\cos(\alpha\pi/2))\}$. Eq. IV.31 is used to obtain starting values for parameters in the nonlinear fit.

The interest of this case study lies in the treatment of the residuals. The results of the multiresponse fit, where instead of f_0 the parameter $\log f_0$ was fitted (because f_0 should be positive), are given in Table IV.1 and the residuals are shown in Fig.IV.2. Note that the real

Table IV.1: Parameter summary for the complex dielectric constant fit

θ	$\underline{\hat{\theta}}$	$\hat{\sigma}_{\underline{\hat{\theta}}}$	Correlation coefficients			
ε_0	4.320	0.006				
ε_∞	2.522	0.018	0.75			
$\log(f_0)$	7.956	0.09	0.46	0.74		
α	0.531	0.010	-0.57	-0.71	-0.93	
β	0.554	0.03	0.53	0.84	0.95	-0.95

residuals show a systematic behaviour, which is related to the way frequencies are set in the experiment. Most probably the actual frequency delivered was affected by the capacity of the sample, and the stepwise change is caused by the decade dial of the oscillator. After a decade correction $f_{\text{corr}} = fK^{[\log(f/100)]}$ with $K = 1.25$ the fit was repeated. Now there appeared to be two outliers in the residuals, which are most clearly visible in the plot of imaginary versus real residuals in Fig.IV.3. After deleting these two cases, the final results are summarized in

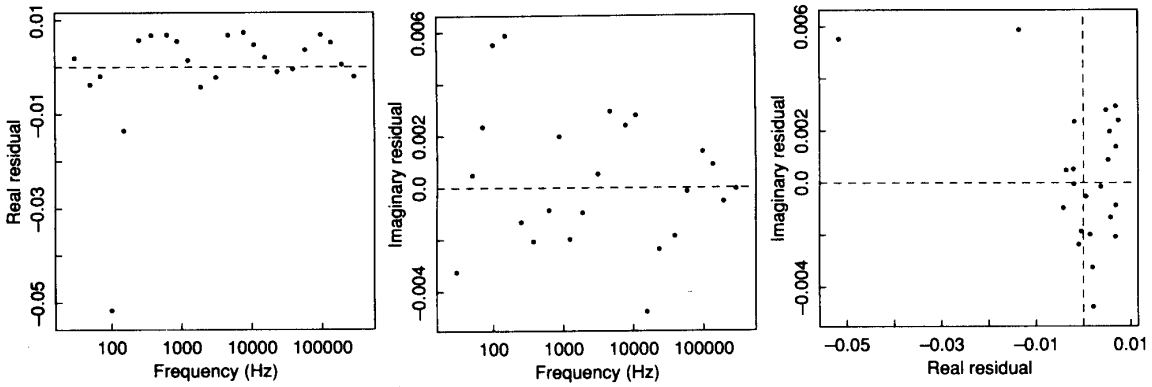

Fig.IV.3. Residuals from the fit to the decade corrected data.

Table IV.2 and the residuals are shown in Fig.IV.4. The residuals from this fit now behaved very well. A note on the difference between the results from Table IV.1 and Table IV.2 is in place here. The appreciable differences in the estimated parameters are due to the decade

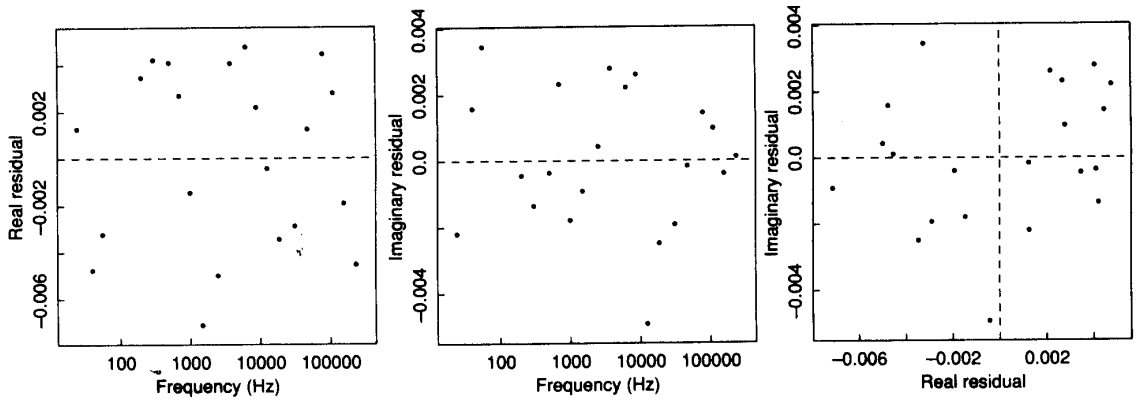

Fig.IV.4. Residuals plus parameter summary of the fit to the corrected and edited data.

Table IV.2: Parameter summary for the complex dielectric constant fit of decade corrected and edited data

θ	$\underline{\hat{\theta}}$	$\underline{\hat{\sigma}}_{\hat{\theta}}$	Correlation coefficients			
ε_0	4.398	0.006				
ε_∞	2.451	0.010	0.53			
$\log(f_0)$	8.25	0.08	0.63	0.91		
α	0.487	0.007	-0.86	-0.75	-0.90	
β	0.57	0.03	0.74	0.91	0.98	-0.95

correction and not to the removal of the two supposes outliers (see Bates and Watts, 1988).

IV.5 Conditionally linear parameters in nonlinear models

Although this section is meant to be quite general, we will illustrate the methods to treat conditionally linear parameters in nonlinear models with an example from global analysis. Suppose the following model describes the time evolution of spectra:

$$\underline{\Psi} = C(\theta)E^T + \underline{\Xi}$$

$$vec(\underline{\Psi}) = vec(C(\theta)E^T) + vec(\underline{\Xi}) = (I_n \otimes C(\theta))vec(E^T) + vec(\underline{\Xi}) \qquad \text{Eq. IV.32}$$

where the $m \times n$ matrix Ψ denotes the measured time resolved spectra, measured at m time instants t_i, and n wavelengths λ_j. Matrix Ξ represents the noise and is, like Ψ, $m \times n$. The matrices C and E, are of dimension $m \times n_{\text{comp}}$ and $n \times n_{\text{comp}}$, respectively. The matrix C

contains n_{comp} columns, each representing the concentration profile of a component, and is described by a kinetic model with parameter vector θ. The matrix *E* also contains n_{comp} columns, each representing the spectral shape of a component. Note that in the vectorial notation we have used Eq. IV.10. For fixed θ Eq. IV.32 represents a multivariate Gauss-Markoff model with design matrix $C(\theta)$ and parameters E^T, which are called the *conditionally* linear parameters. Note that in the special case that $n = 1$ we have an ordinary nonlinear, weighted sum model with conditionally linear amplitude parameters. The model of Eq. IV.32 is called a *partially linear* or *separable nonlinear* model.

There are special computational methods which exploit this partial linearity.

For fixed θ the solution of Eq. IV.32 is given by:

$$\hat{\underline{E}}^T(\theta) = C^{\dagger}(\theta)\underline{\Psi} \qquad \text{Eq. IV.33}$$

$C^{\dagger}$ denotes the Moore-Penrose generalized inverse of *C*, which is of full rank. Note that the following is completely analogous to §II.11, *Determining the linear least squares estimates using the QR decomposition* , except for the dependence upon the parameters θ. Therefore the method is called *variable projection* (Golub and Pereyra, 1973).

Following Kaufman (1975) and Golub and Leveque (1979) we perform a *QR* decomposition of $C(\theta)$:

$$C(\theta) = \begin{bmatrix} Q_1(\theta) & Q_2(\theta) \end{bmatrix} \begin{bmatrix} R(\theta) \\ 0 \end{bmatrix} = Q_1(\theta)R(\theta) \qquad \text{Eq. IV.34}$$

where $Q_1(\theta)$ and $Q_2(\theta)$ are, respectively, $m \times n_{comp}$ and $m \times (m - n_{comp})$ matrices which together form the orthogonal matrix $Q(\theta)$. $R(\theta)$ represents an $n_{comp} \times n_{comp}$ upper triangular matrix.

Combining Eq. IV.33 and Eq. IV.34 we have:

$$\hat{\underline{E}}^T(\theta) = R^{-1}(\theta)Q_1^T(\theta)\underline{\Psi} \qquad \text{Eq. IV.35}$$

Combining Eq. IV.34 and Eq. IV.35 we find for the residual matrix *Z*:

$$\underline{Z}(\theta) = \underline{\Psi} - C(\theta)\hat{\underline{E}}^T(\theta) = (I - Q_1(\theta)Q_1^T(\theta))\underline{\Psi} = Q_2(\theta)Q_2^T(\theta)\underline{\Psi} \qquad \text{Eq. IV.36}$$

where $Q_2(\theta)Q_2^T(\theta)$ is an orthogonal projection matrix.

We distinguish two measurement situations:

i] independent and identically distributed observations at time instants t_i of responses at wavelengths λ_j:

$$D(\underline{\xi}_{t_i\lambda_j}, \underline{\xi}_{t_k\lambda_l}) = \delta_{ik}\sigma_{jl} \qquad \text{Eq. IV.37}$$

where the Kronecker delta δ_{ik} indicates that different observations are independent, and σ_{jl} is an element of the response covariance matrix Σ. Σ describes the variances of and correlations between responses measured simultaneously at different wavelengths, and is independent of the time instant t_i. The covariance matrix of the vector representation of Ξ is the Kronecker product of Σ with I_m:

$$D(\mathrm{vec}\underline{\Xi}) = \Sigma \otimes I_m \qquad \text{Eq. IV.38}$$

The determinant to be minimized (Eq. IV.25) as a function of the unknown parameters θ is given by

$$\underline{\Upsilon}(\theta) = \det(\underline{Z}^T(\theta)\underline{Z}(\theta)) = \det(\underline{\Psi}^T Q_2(\theta)Q_2^T(\theta)\underline{\Psi}) \qquad \text{Eq. IV.39}$$

since $Q_2^T(\theta)Q_2(\theta) = I_{m-n_{\mathrm{comp}}}$. Minimization of $\underline{\Upsilon}(\theta)$ gives us the maximum likelihood estimate $\underline{\hat{\theta}}$. Thus we have eliminated the linear parameters *E* and have arrived at a residual matrix $Q_2^T(\theta)\underline{\Psi}$ with $m - n_{\mathrm{comp}}$ observations of *n* responses. Prerequisites for determinant minimization are that the number of observations n_{obs} should be greater than the number of unknown parameters n_{par} and not smaller than the number of responses n_{resp}:

$$n_{\mathrm{obs}} \geq n_{\mathrm{resp}} \qquad n_{\mathrm{obs}} > n_{\mathrm{par}} \qquad \text{Eq. IV.40}$$

This leaves as degrees of freedom: $\mathrm{df} = m - n_{\mathrm{comp}} - n_{\mathrm{par}}$, where $n_{\mathrm{par}} = \dim(\theta)$. Thus when $\mathrm{df} \leq 0$ or when $n_{\mathrm{obs}} = m - n_{\mathrm{comp}} < n = n_{\mathrm{resp}}$ (Eq. IV.40) the multiresponse problem is unsolvable. We must then resort to the next situation.

ii] independent and identically distributed observations at wavelengths λ_i and time instants t_j. Thus the variance ς^2 is assumed to be independent of both t_i and λ_j. We now have

$$\mathrm{cov}(\underline{\xi}_{t_i\lambda_j}, \underline{\xi}_{t_k\lambda_l}) = \varsigma^2\delta_{ik}\delta_{jl} \qquad \text{Eq. IV.41}$$

When the noise influences $\underline{\xi}_{t_i\lambda_j}$ are independent and identically distributed, across wavelength and time, the nonlinear least squares estimator is the maximum likelihood estimator, and thus the best estimator. The sum of squares to be minimized as a function of the unknown parameters θ is given by

$$\underline{\Upsilon}(\theta) = \text{trace}(\underline{Z}^T(\theta)\underline{Z}(\theta)) = \text{trace}(\underline{\Psi}^T Q_2(\theta)Q_2^T(\theta)\underline{\Psi}) \qquad \text{Eq. IV.42}$$

The number of observations now is $(m - n_{\text{comp}}) \times n$ which leaves as degrees of freedom: $\text{df} = (m - n_{\text{comp}}) \times n - n_{\text{par}}$. In the highly improbable case that this $\text{df} \leq 0$ the problem cannot be solved.

*IV.5.1 Implementation of the variable projection method**

There are some implementational details about the actual minimization of Eq. IV.39 or Eq. IV.42. To begin with, most minimization methods (see §III.2) require a gradient of the residuals. We can use the so-called Kaufman approximation. We require the derivatives $\frac{\partial}{\partial\theta}Q_2^T(\theta)\underline{\Psi}$. Following again Kaufman (1975) and Golub and Leveque (1979) we recall from Eq. IV.34 that $Q_2^T(\theta)C(\theta) = 0$ and thus

$$0 = \frac{\partial}{\partial\theta}(Q_2^T(\theta)C(\theta)) = \left(\frac{\partial}{\partial\theta}Q_2^T(\theta)\right)C(\theta) + Q_2^T(\theta)\left(\frac{\partial}{\partial\theta}C(\theta)\right) \qquad \text{Eq. IV.43}$$

This leads to the approximation $\frac{\partial}{\partial\theta}Q_2^T(\theta) \approx -Q_2^T(\theta)\left(\frac{\partial}{\partial\theta}C(\theta)\right)(C(\theta))^\dagger$ which gives, using Eq. IV.33

$$\frac{\partial}{\partial\theta}Q_2^T(\theta)\underline{\Psi} \approx -Q_2^T(\theta)\left(\frac{\partial}{\partial\theta}C(\theta)\right)\underline{\hat{E}}^T(\theta) \qquad \text{Eq. IV.44}$$

Now some remarks on dimensions of typical problems are in order. With some types of measurements the matrix of conditionally linear parameters can have dimensions of say $m \times n_{comp} = 500 \times 5$, whereas θ may be of dimension 10. Thus the variable projection method allows to solve implicitly for 2500 of the 2510 parameters. Next only a gradient $\frac{\partial}{\partial\theta}C(\theta)$ needs to be calculated and stored. It goes without saying that with ordinary methods such a problem would have been virtually unsolvable.

*IV.5.2 Covariance matrix of conditionally linear parameters**

Since by using the variable projection method the criterion does no longer explicitly depend upon the conditionally linear parameters, we can derive only linear approximation summary statistics for the nonlinear parameters θ (using Eq. III.20, or Eq. IV.27). To find a linear approximation covariance matrix for the conditionally linear parameters we must consider the full parameter estimation problem.

For the nonlinear least squares case we estimate the covariance matrix

$$D\begin{pmatrix} \hat{\underline{\theta}} \\ \mathrm{vec}(\hat{\underline{E}}^T) \end{pmatrix} = \hat{\underline{\sigma}}^2 (J^T J)^{-1}_{\theta = \hat{\theta},\, E = \hat{E}} \qquad \text{Eq. IV.45}$$

where J denotes the $mn \times (n_{\mathrm{par}} + n \cdot n_{\mathrm{comp}})$ Jacobian matrix

$$J = \frac{\partial(\mathrm{vec}(CE^T))}{\partial(\theta, \mathrm{vec}E^T)} = \left[(E \otimes I_m) \frac{\partial(\mathrm{vec}C)}{\partial\theta} \; I_n \otimes C \right] = \left[G \; I_n \otimes Q_1 R \right] \qquad \text{Eq. IV.46}$$

where we have abbreviated the gradient $(E \otimes I_m)\frac{\partial(\mathrm{vec}C)}{\partial\theta} \equiv \mathrm{vec}(\dot{C}E^T) \equiv G$. After some algebra (using the matrix inversion lemma, Eq. A.4, for the block matrix $J^T J$) we find the following covariance matrices (all to be evaluated at $\theta = \hat{\theta}, E = \hat{E}$):

$$D(\hat{\underline{\theta}}) = \hat{\underline{\sigma}}^2 P^{-1} \qquad \text{Eq. IV.47}$$

where $P = G^T (I_n \otimes Q_2 Q_2^T) G = \left(\frac{\partial(\mathrm{vec}C)}{\partial\theta}\right)^T (E^T E \otimes Q_2 Q_2^T) \frac{\partial(\mathrm{vec}C)}{\partial\theta}$ and

$$\begin{aligned} D(\mathrm{vec}(\hat{E}^T)) &= \hat{\underline{\sigma}}^2 (I_n \otimes R^{-1}R^{-T} + (I_n \otimes R^{-1}Q_1^T) G P^{-1} G^T (I_n \otimes Q_1 R^{-T})) \\ &= \hat{\underline{\sigma}}^2 \left(I_n \otimes R^{-1}R^{-T} + (E \otimes R^{-1}Q_1^T) \frac{\partial(\mathrm{vec}C)}{\partial\theta} P^{-1} \left(\frac{\partial(\mathrm{vec}C)}{\partial\theta}\right)^T (E^T \otimes Q_1 R^{-T}) \right) \end{aligned} \qquad \text{Eq. IV.48}$$

For the spectral parameters the estimated covariance depends upon the wavelength because of the second term in Eq. IV.48. This is in contrast to the Gauss-Markoff model with a priori known C. For practical purposes only Eq. IV.48 is used, in particular the second form which can be calculated jointly with Eq. IV.44 (using e.g. IMSL routine DLQRSL).