Interaction Terms in Non-linear PLS via Additive Spline Transformation

Durand J.F.[1] and Lombardo R.[2]

[1] Laboratoire de Probabilités et Statistique,
University of MontpelierII, France
E-Mail: jfd@helios.ensam.inra.fr

[2] Department of Accounting and Quantitative Methods,
II University of Naples, Italy.
E-Mail: rosaria.lombardo@unina2.it

Preprint of the paper in "Between Data Science and Applied Data Analysis",
Studies in Classification, Data Analysis, and Knowledge Organization.
Editors M.Schader, W.Gaul and M.Vichi, Springer, 2003, 22-29.

Abstract. This paper aims to extend the non-linear additive Partial Least Squares model via Spline transformation (PLSS,Durand 2001) to the detection and introduction of significant interaction terms as linear manifolds of multivariately transformed variables (Lombardo & van Rijckevorsel 2000).

Keywords: PLS via Spline regression, Interaction terms, Multivariate B-spline, PRESS criterion.

1 Introduction

In a typical Partial Least Squares regression (Wold 1966, Tenenhaus 1998) one assumes that there exists an underlying linear relationship between the response and predictor variables. Sometimes there is reason to doubt this assumption and a non-linear transformation of the variables might be useful to reveal the model underlying the data, producing a linear relationship among the transformed variables. This paper introduces a generalization of PLSS, the non-linear additive Partial Least Squares model via spline transformation (Durand 2001) by the inclusion of variable interactions. Additive splines for PLS regression are based on the transformation of each predictor by B-spline functions. In this paper, the additive model involves both univariate and multivariate B-spline transformations. The use of interaction terms in this new version of PLSS means looking for functions of two or more variables which help to understand a complex and or noisy relationship. The interaction degree depends on the number of variables involved in the analysis. In this first exercise in order not to heavy the computational burden of the procedure, we focus on the bivariate interactions. The usual strategy to select the optimal interaction terms (MARS, Friedman 1991) implies a forward and a backward phase: to let the forward phase produce a complete number of predictors and then have the backward one trim the model back to an appropriate size. In the PLS component based regression context, taking into account the whole set of main effects and interactions terms can theoretically be envisaged. However, the price to be payed by PLSS for interactions through tensor products of B-spline functions which is tantamount to expending the column dimension of the new design matrix, is excessive when the method is applied on typical PLS number of predictors. A selection of interaction candidates is proposed that is based on a criterion depending on both goodness-of-fit and goodness-of-prediction.

Piecewise polynomials or splines extend the advantages of polynomials to include flexibility and local effect of parameter changes. Due to local B-spline basis functions, the PLSS model is resistant to the presence of extreme values of the predictors. Note that locating knots in empty regions is well accepted by PLSS since regressions are made on uncorrelated components that are linear compromises of centered univariate and multivariate B-spline functions.

In the second section we briefly reviewed PLSS, while the detection and importance of interaction terms is illustrated in section 3. The details of the computational procedure are presented in section 4. In the last section a real example illustrates the capabilities of PLS on a sensory data set in the presence of outliers and non-linear interactions.

2 Modelling main effects by Partial Least Squares through Splines

Let $\mathbf{Y} = [Y^1|..|Y^q]$ be the quantitative or categorical $n \times q$ response matrix and $\mathbf{X} = [X^1|..|X^p]$ the $n \times p$ observation matrix of the predictors on the same n statistical units. Various articles have been presented in literature to generalize PLS in the non-linear framework, most of them replace the standard regression with non linear functions (Wold, Kettaneh-Wold, Skagerberg 1989, Durand & Sabatier 1997). In this paper we enhance a recently devised non-linear additive model for PLS via B-spline transformations that is called PLSS (Durand 2001), so introducing a non parametric character in the PLS regression. In short, PLSS is defined as the usual linear PLS regression of $\mathbf{Y}$ onto the space spanned by the centered coding matrix $\mathbf{B} = [\mathbf{B}^1|..|\mathbf{B}^p]$ ($n \times r$ matrix), whose columns are constructed by transforming the predictors through B-spline basis functions (De Boor 1978)

$$PLSS(\mathbf{X}, \mathbf{Y}) = PLS(\mathbf{B}, \mathbf{Y})\,.$$

Notice that Principal Component Analysis (PCA) can be seen as a "self-PLS" regression of $\mathbf{X}$ onto itself. Then, in the same way as a particular property, comparing PLSS with Non-Linear Principal Component Analysis (NLPCA, Gifi 1990), we can say that NLPCA can be considered as a "self-PLS" of the coding matrix $\mathbf{B}$ onto itself

$$PLSS(\mathbf{X}, \mathbf{Y} = \mathbf{B}) = PLS(\mathbf{B}, \mathbf{Y} = \mathbf{B}) = PCA(\mathbf{B}) = NLPCA(\mathbf{X})\,.$$

In the multivariate context, the column dimension of the column centered coding matrix $\mathbf{B}$ is the sum $r = \sum_{j=1}^{p} r_j$, where r_j, the column dimension of the block $\mathbf{B}^j$, is given by $r_j = d_j + K_j$ if d_j denotes the degree of the local polynomial and K_j the number of interior knots. Using B-spline functions, we can simultaneously treat variables of different nature (qualitative and quantitative) and at the same time we warrant against the presence of extreme values of the predictors since B-spline functions have a local support. A multi-response additive spline model is a fit of the form

$$\hat{Y}_A^i = \sum_{j=1}^{p} s_A^{i,j}(X^j) = \sum_{j=1}^{p} \mathbf{B}^j \hat{\beta}_A^{i,j} \qquad (\text{for} \quad i = 1, .., q)$$

where the spline function $s_A^{i,j}$ measures the additive influence of the predictor X^j on the response Y^i and depends on the number A of the carried out components. This is an univariate spline function summing over all single variable basis functions involving only x_j. When the number A is equal to the rank of $\mathbf{B}$, then PLSS is identical to the usual Least-Squares Splines estimator, see Durand (2000). The risk of overfitting related to an increasing column dimension for the new design matrix $\mathbf{B}$ is well supported by PLSS which inherits the advantages of the standard PLS method.

The non-linear additive influence of the predictors on the i^{th} response is interpreted by looking at the most significant coordinate function plots $(x^j, s_A^{i,j}(x^j))$. In practice, the selection of predictors can be reached by means of different strategies (Durand 2000): ordering in decreasing order the range of $s_A^{i,j}(X^j)$ for $j = 1, \ldots, p$, or grouping the predictors with the same coordinate shape.

The building-model stage consists of finding a balance between "goodness" (of fit and prediction) and "thriftiness" (of dimension for both the number A of PLSS components and r the total dimension). In order to evaluate the goodness-of-fit, we look at the criterion $R^2(A)$, that is the proportion of the total $\mathbf{Y}$ variance accounted for by the PLSS components $t^1, .., t^A$:

$$R^2(A) = \frac{1}{q} \sum_{i=1}^{q} R^2(Y^i, span(t^1, .., t^A))$$

which is an increasing function of A. In the same way the construction of the $PRESS^i(A)$ criterion, identical to that of the usual PLS regression, is computed for each response, while the total $PRESS(A) = \sum_{i=1}^{q} PRESS^i(A)$ is a function of A expected to be firstly decreasing. In conclusion to avoid overfitting problems, we look for parsimonious models with the best values of both $R^2(A)$ and $PRESS(A)$ criteria.

3 Interaction terms in non-linear Partial Least Squares

The problem to replace complex and/or noisy relationships between variables in order to reveal the model underlying the data is a vast topic in literature. To generalize PLS, Hoskuldsson (1988) discussed a model with quadratic and interaction terms in the original variables. In the present paper we propose to include interaction terms not involving the original variables but the non-linearly transformed ones. Suppose we have three chemical substances, two of them react into a new polluting hazardous substance correlated to the third one. The coordinate-wise product of two of the variables is correlated, our problem is to identify the best interaction terms out of all possible interactions, avoiding multicollinearity problems.

The higher order interactions in non-linear PLS allow to detect dependencies i.e. interactions among variables, using multivariate transformations. The use of interaction terms means looking for functions of two or more variables. The order number of interaction depends on the number of variables involved in the analysis. In this paper the adopted functions are bivariate splines, which are tensor products of B-spline univariate functions.

In order to investigate more complex interrelationships, the PLSS fit is casted in the following ANOVA decomposition which is the sum of the main effects and the accepted bivariate interactions

$$\hat{Y}^i_A = \sum_{j=1}^{p} s_A^{i,j}(X^j) + \sum_{(j,j')\in\mathcal{I}} s_A^{i,jj'}(X^j, X^{j'}) \tag{1}$$

where $\mathcal{I}$ is the set of the accepted couples of interactions. A bivariate ANOVA interaction $s_A^{i,jj'}(X^j, X^{j'})$ is a linear compromise of $\mathbf{B}^{jj'}$, the $n \times (r_j r_{j'})$ centered matrix of the tensor product of columns in both $\mathbf{B}^j$ and $\mathbf{B}^{j'}$, respectively. It represents the joint bivariate contribution of X^j and $X^{j'}$ to the model. This model is obtained by the PLS regression of the standardized responses Y on centered spline transformations $\mathbf{B} = [\mathbf{B}^1 \ldots \mathbf{B}^p | \ldots \mathbf{B}^{jj'} \ldots]$. The computational cost is clearly related to the increase of the column dimension for the new design matrix $\mathbf{B}$, which depends on the dimension of the spline space for the predictors and on the detected interactions.

The tuning parameters of the model are now: 1) the degree, number and location knots, 2) the number of components, 3) the interactions accepted by the building-model stage.

Let us firstly detail the first phase of the selection based on the individual evaluation of all possible bivariate interactions. When A is fixed, the criterion for accepting one candidate interaction is based on the gain in fit and prediction compared to that from the model with only the main effects. Denoting "m", and "m+i", the models including respectively the main effects and the main effects plus one interaction, then, the criterion used to evaluate that interaction, i, is defined by

$$CRIT(A) = \frac{R^2_{m+i}(A) - R^2_m(A)}{R^2_m(A)} + \frac{PRESS_m(A) - PRESS_{m+i}(A)}{PRESS_m(A)}$$

Note that the second term in the sum is generally larger than the first when $R^2_m(A)$ is close to one, that is when the main effects fit is good. One interaction is accepted if $CRIT > 0$, thus providing a natural selection procedure to decide on how many candidate interactions, among all the $(p-1)p/2$ possible ones, are accepted. Then, the selected interactions are decreasingly ordered to add them step-by-step to the main effects model.

4 The building-model stage

In this section we illustrate the main aspects of the computational procedure used to compute and select the best interaction terms in the non-linear model.

0 Preliminary phase: the main effects model.
Decide on the spline parameters as well as on A for the main effects model (m): denote $PRESS_m(A)$ and $R^2_m(A)$.

1 Phase 1: individual evaluation of candidate interactions.
Each interaction i is separately added to the main effects. Eliminate interactions such that $CRIT(A) < 0$. Then, order decreasingly the accepted candidate interactions.

2 Phase 2: stepwise forward building-model stage.
Set $PRESS_0 = PRESS_m(A)$, $i = 0$ and index by i the partial current model. Repeat the procedure: add to the model i the accepted interaction candidate $i+1$ only if the corresponding optimal PRESS fulfils the condition $PRESS \leq PRESS_i$.

5 Application: the orange juice data

The example from sensometrics herein outlined, is discussed in Durand (2001). The real data set consists of 10 explanatory variables illustrating the mineralogical properties of 24 orange juices and one sensory response. For confidentiality, the 24 orange juices have been named by the capital letters "A",..,"X". We assume that they all have identical quantity of fruit and same glucid content. The predictors are given by eight mineralogical characters SiO_2, Na, K, Ca, Mg, Cl, SO_4, HCO_3 and their sum Sum, plus the conductivity $COND$. The response is given by the sensory variable $HEAVY$, whose protocol of measurements is not revealed for confidentiality.
The a priori knowledge of this sensory data permits to select the degree 2 and locate knots for each variable at the points specified in Table 1.

Table 1: Selected knots for the predictors

COND	SiO_2	**Na**	**K**	**Ca**	**Mg**	**Cl**	SO_4	HCO_3	**Sum**
400	10	10	2.5	160	40	4	400	100	600
1600	20	40	5	400	110	11	1700	300	2600
	40					30		500	

Figure 1 displays the comparison of PRESS values between the main effects model and the final model with interactions, both computed with 2 dimensions. The respective R^2 (from 0.917 to 0.93) measures the gain in goodness-of-fit when the interaction terms are included as predictors into the model.

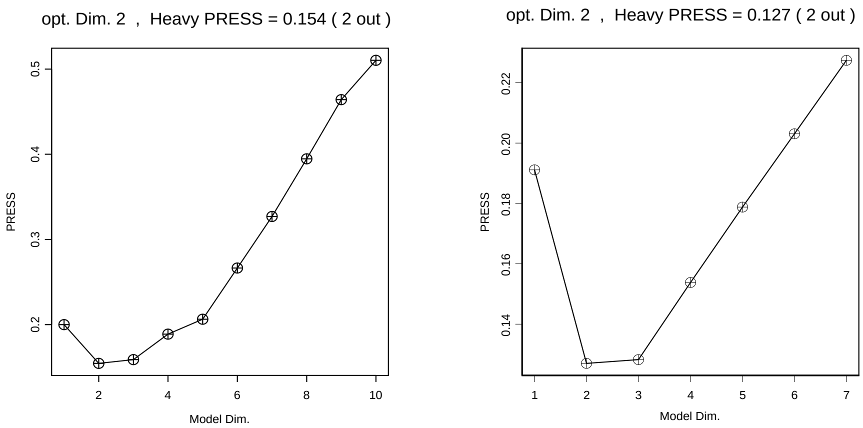

Figure 1: PRESS plots: main effects model (left), interactions model (right)

Figure 2, left-side, displays the phase 1 individual evaluation and selection of the 45 possible interactions. The right-side of this figure shows the PRESS values of the successive steps in the phase 2 building-model-stage. The final model includes 10 interactions added to the 10 main effects.

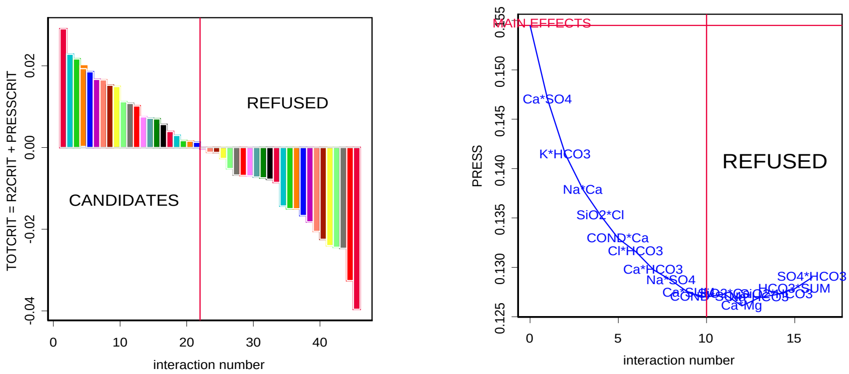

Figure 2: Phase 1 (left) and phase 2 (right) of the building-model stage

In Figure 3, univariate and bivariate ANOVA functions have been ordered according to their decreasing additive influence on the response $HEAVY$: the graphs are ordered from left to right and up to down according to the range of the transformed variables. Vertical in the main effects plots indicate the location of the knots. We notice that the most influent interaction $Ca * SO_4$ belongs to the group of the five largest predictors. Furthermore, one can see that extreme values of the predictors (for example, orange juice P in the variables SO_4, Ca, Sum) do not perturb the coordinate functions due to the location of knots.

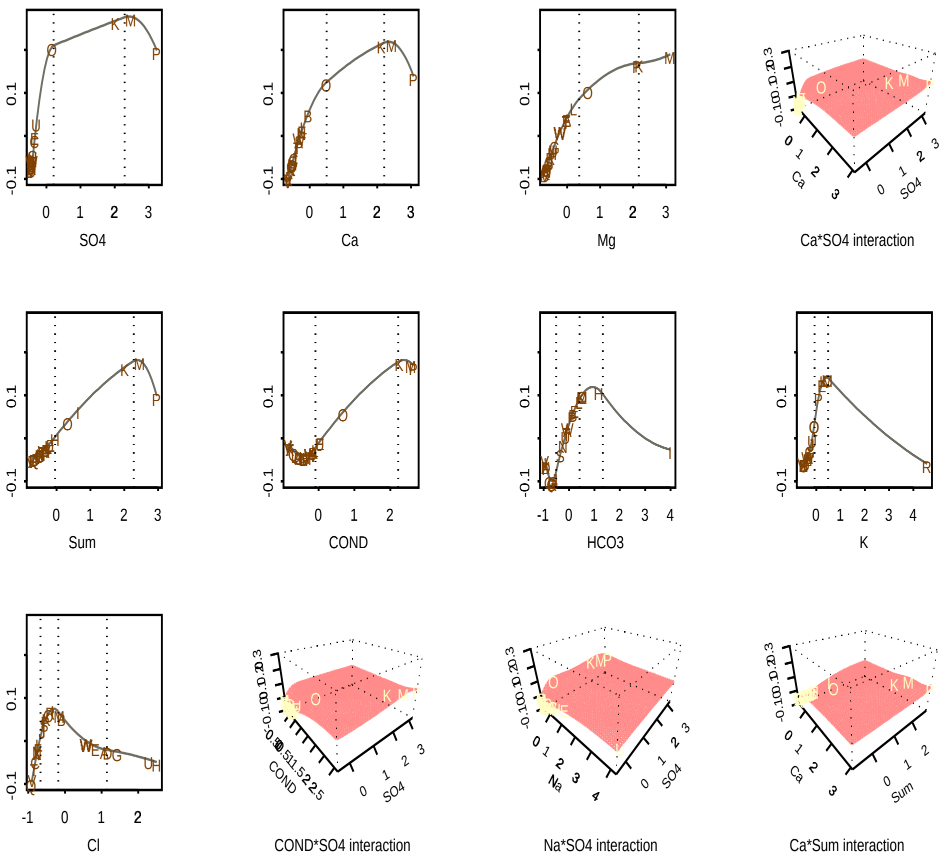

Figure 3: ANOVA functions of main influence on HEAVY (2 dimensions).

6 Conclusion

PLSS with interactions is a straightforward extension of *PLSS* (Durand 2000, 2001), it inherits from *PLSS* the robustness against the presence of extreme values of predictors and from *PLS* the traditional robustness against both scarcity of data and multicollinearity of the variables. The method does not automatically estimate the optimal transformations of univariate and bivariate predictors, but exploring different sets of tuning spline parameters allows the user to experiment with a wide range of PLS tools from linearity to non-linearity through local polynomial models with interactions.

References

De Boor C. (1978) A Practical Guide to Splines, Springer-Verlag, Berlin, 1978.

Durand J.F. (2000) La Régression Partial Least Squares Spline, PLSS, guide d'utilisation sous S-plus. Rapport de Recherche 00-06, Groupe de Biostatistique et d'Analyse des Systèmes ENSAM-INRA-UM II.

Durand J.F. (2001) Local Polynomial additive regression through PLS and Splines: PLSS. In Chemometrics & Intelligent Laboratory Systems, 58, 235-246.

Durand J.F. and Sabatier R. (1997) Additive Splines for Partial Least Squares Regression. In Journal of the American Statistical Association, vol.92, n.440.

Friedman, J. (1991) Multivariate Adaptive Regression Splines. The Annals of Statistics, 19,1:1-141.

Gifi, A.(1990) Non-linear Multivariate Analysis, DSWO Press.

Hoskuldsson P. (1988) PLS Regression Methods. Journal of Chemometrics, 2,211-228.

Lombardo, R. Rijckevorsel, J.van (2000): Interaction Terms in Homogeneity Analysis: Higher order non-linear Multiple Correspondence Analysis. In "Studies in Classification, Data Analysis and knowledge Organization" Springer.

Tenenhaus, M. (1998): La Régression PLS, Théorie et Pratique. Editions Technip, Paris.

Wold, H. (1966): Estimation of principal components and related models by iterative least squares. In Multivariate Analysis, (Eds.) P.R. Krishnaiah, New York: Academic Press, 391-420.

Wold, H. (1992): Non Linear Partial Least Squares Modeling II. Spline Inner Relation. In Chemometrics and Intelligent laboratory Systems, 14, 71-84.

Wold, S., Kettaneh-Wold, H., Skagerberg B.(1989): Non Linear Partial Least Squares Modeling. In Chemometrics & Intelligent Laboratory Systems, 7,53-65.