

THE COLLINEARITY PROBLEM IN LINEAR REGRESSION. THE PARTIAL LEAST SQUARES (PLS) APPROACH TO GENERALIZED INVERSES*

S. WOLD†, A. RUHE‡, H. WOLD§ AND W. J. DUNN III¶

Abstract. The use of partial least squares (PLS) for handling collinearities among the independent variables X in multiple regression is discussed. Consecutive estimates (rank 1, 2, \dots) are obtained using the residuals from previous rank as a new dependent variable y . The PLS method is equivalent to the conjugate gradient method used in Numerical Analysis for related problems.

To estimate the "optimal" rank, cross validation is used. Jackknife estimates of the standard errors are thereby obtained with no extra computation.

The PLS method is compared with ridge regression and principal components regression on a chemical example of modelling the relation between the measured biological activity and variables describing the chemical structure of a set of substituted phenethylamines.

Key words. collinearity, linear regression, conjugate gradients, principal components, cross validation, chemometrics

1. Introduction. In multiple regression, linear or nonlinear, collinearities among the independent variables x_j sometimes cause severe problems. (For notation, see below equation (1)). The estimated coefficients $\hat{\beta}_j$ can be very unstable, and thereby far from their target values. In particular, this makes predictions by the regression model to be poor.

In many chemical applications of multiple regression, like the present example of relationships between chemical structure and biological activity, the predictive properties of the models are of prime importance and the regression estimates therefore often need to be stabilized. The present example can be seen as a special case of response surface modelling, an area where the collinearity problem has been recognized as a serious problem (Box and Draper (1971), Gorman and Toman (1966), Draper and Smith (1966)).

In applied work, the collinearity problem is often handled by selecting a subset of variables by a stepwise procedure. See Hocking (1976) for a review. We shall not consider this subset strategy here, but shall limit ourselves to the data analysis with all variables included in the model.

2. Existing methods. Three principal ways are described in the statistical literature to accomplish a stabilization of the regression estimates $\hat{\beta}_j$ in a given regression model. Adopting the usual notation (see e.g. Draper and Smith (1981, p. 72)), the linear model is:

$$(1) \quad y = Xb + e.$$

Here X is a $n \times p$ matrix, containing the values of the p predictor variables at the n data points, $b = (\beta_1, \dots, \beta_p)'$ is a p -dimensional column vector containing the regression coefficients, and $e = (\varepsilon_1, \dots, \varepsilon_n)'$ is an n vector containing the errors which

* Received by the editors April 19, 1982, and in revised form April 28, 1983.

† Research Group for Chemometrics, Institute of Chemistry, Umeå University, S-901 87 Umeå, Sweden.

‡ Present address: Department of Computer Science, Chalmers University of Technology, S-41296 Göteborg, Sweden.

§ Department of Statistics, Uppsala University, S-751 21 Uppsala, Sweden.

¶ Department of Medicinal Chemistry, College of Pharmacy, University of Illinois at the Medical Center, Chicago, Illinois 60680.

are assumed to be uncorrelated, to be normal and to have the same variance. The prime ' denotes the transpose of a vector or matrix.

The first way, usually called *ridge regression* (Hoerl and Kennard (1970)) is based on adding a small number, k , to all elements in the diagonal of the moment matrix $X'X$. Thus, instead of inverting $X'X$, one inverts $(X'X + k \cdot I)$, which gives the regression estimates

$$(2) \quad b_{\text{ridge}} = (X'X + k \cdot I)^{-1} \cdot X'y.$$

The "ridge parameter," k , is usually chosen between 0 and 1. The specific value of k is often based on an inspection of a plot of the regression estimates against k (see Fig. 1 for an example). Golub, Heath and Wahba (1979) estimate k directly from the data using generalized cross validation. Ridge regression has recently been reviewed by Draper and Van Nostrand (1979) and by Hocking (1976) (see also Smith and Campbell (1980)). The solution of nonlinear regression using ridge estimates in the iterative updating is a well established numerical practice (see Marquardt (1970) where the correspondence between the two situations is discussed).

The second approach to the stabilization of the regression estimates is based on the contraction of $X'X$ to a matrix of smaller rank. This is, for instance, accomplished by expanding $X'X$ in terms of its eigenvectors (principal components) and then retaining only the first r of these p eigenvectors to represent $X'X$ ($r < p$). This gives the solution the form of a generalized inverse (Marquardt, 1970). A more elegant and more numerically stable formulation of that approach is based on the singular value decomposition (SVD, see Golub and Kahan (1965)) of X itself. The r first singular vectors can then be regarded as new independent predictor variables, *principal components regression* (see Hocking (1976) for a review).

Marquardt (1970) and others (Hawkins (1975), Mayer and Willke (1973)), showed the close similarity (but nonequivalence) between ridge and generalized inverse estimates. A large literature on closely related problems in numerical analysis, so-called ill-posed problems, addresses the collinearity problem in linear models along essentially the same two lines (see Varah (1979), Björck and Elden (1979), Wahba (1977)).

A third, less often used approach, is the so called *James-Stein estimates*. These consist of the ordinary least squares estimates multiplied by a shrinking factor φ ($0 < \varphi < 1$). The use of these shrunk estimates in multiple regression has recently been reviewed by Draper and van Nostrand (1979) and Hocking (1976).

3. The PLS method. In the present article we investigate the properties of so called PLS estimates (partial least squares) adapted to the multiple regression problem. PLS was recently developed for modelling information-scarce situations in Social Sciences (H. Wold, 1975, 1982). We show that the present PLS estimation is a variant of the conjugate gradient method developed in numerical analysis for the calculation of generalized inverses (Hestenes and Stiefel (1952), Golub and Kahan (1965), Paige and Saunders (1982)). Each step of the conjugate gradient algorithm gives the PLS estimate of the corresponding rank.

The predictive significance of each sequential estimate can be tested by cross validation with only a small amount of additional computation. One simultaneously obtains jackknife estimates of the regression coefficient standard errors. Thus an efficient stopping rule is obtained; the rank is used which gives the model the best predictive properties in the cross-validatory sense.

The method is particularly attractive because (1) only two vector-matrix multiplications are needed for each successive rank estimate; and (2) the calculations are

performed on the raw data y and X and the moment matrix needs not be calculated. This makes the method suitable for large problems where the amount of computation with standard methods of matrix inversion or diagonalization becomes prohibitive. Examples of such large problems are found in protein X-ray crystallography (Konnert (1976)) and geodesy (Kolata (1978)). With the proliferation of microcomputers, fast and simple-to-program methods are of interest also for small and moderate problems.

4. Details of the estimation. The PLS algorithm, in its general, mode A formulation, deals with variables blocked in q blocks, and forms a sequence of rank one approximations to the combined data matrix. In this paper, we consider only the case with two blocks, one of them furthermore restricted to consisting of only one variable. Let the data matrices for the two blocks be X and y , and denote the combined matrix by

$$Z = [X|y].$$

We then successively form a sequence of residual matrices Z_s , using the following algorithm:

ALGORITHM PLS.

1. Start $Z_1 = [X|y]$, $b_0 = 0$.
2. For $s = 1, 2, \dots$, until $\|Z_s\|$ is small
 1. $u_s = X_s X'_s y_s / \|X'_s y_s\|$.
 2. $c_s = Z'_s u_s / u'_s u_s$, $c_s = (a'_s, \rho_s)'$.
 3. $Z_{s+1} = Z_s - u_s c'_s$.
 4. Solve $A'_s b_s = r_s$, $r_s = (\rho_1, \dots, \rho_s)'$ for b_s .

We first note that

$$U_r = [u_1, \dots, u_r]$$

builds up an orthogonal basis of the range of X , since X_{s+1} is the projection of X orthogonal to U_s , (see step 2.3):

$$X_{s+1} = X_s - \frac{u_s u'_s X_s}{u'_s u_s} = \left(I - \frac{u_s u'_s}{u'_s u_s} \right) X_s = \left(I - \frac{u_s u'_s}{u'_s u_s} \right) \cdots \left(I - \frac{u_1 u'_1}{u'_1 u_1} \right) X$$

making $u'_{s+1} U_s = 0$. Noting that step 2.1 is a gradient step, we see then that the PLS algorithm is actually equivalent to a conjugate gradient algorithm applied to the normal equations of the system (1). The right vectors denoted by a_s are not orthogonal, so in order to solve for the regression coefficients, we have to update the solution of an underdetermined system in step (2.4).

The formulation in algorithm PLS is of interest since it shows the close connection to the principal components regression and total least squares (Golub, van Loan (1980)) algorithms, in that they all build up Z as a sum of rank one matrices. However for actual computation, especially for large problems where the updating step (2.3) becomes time-consuming, we suggest that a reliable implementation of conjugate gradients be used. The algorithm LSQR (Paige and Saunders (1982)) is the preferred choice.

ALGORITHM LSQR. (see Paige and Saunders (1982) for details).

1. Start $\theta_1 v_1 = X' y$, $\rho_1 p_1 = X v_1$.
2. For $s = 1, 2, \dots$
 1. $\theta_{s+1} v_{s+1} = X' p_s - \rho_s v_s$.
 2. $\rho_{s+1} p_{s+1} = X v_{s+1} - \theta_{s+1} p_s$.

The positive numbers θ_s and ρ_s are normalization coefficients chosen to give the vectors v_s and p_s unit length. We have given the algorithm bidiag 2 of Paige and Saunders (1982), which is a simple transformation of the algorithm bidiag 1 that they have actually implemented.

We see that the vectors p_s of algorithm LSQR are proportional to the vectors u_s of algorithm PLS. In fact, the two algorithms are equivalent, and give the same sequence of solutions. Convergence occurs when θ_{s+1} or ρ_{s+1} become negligible, and since the vectors are orthogonal, this occurs for $s=r$, the rank of X , if not earlier. Often sufficiently good numerical results are obtained much earlier.

5. Data scaling. The PLS estimates depend on the scaling of the variables x_j as do ridge and principal components regression estimates. Thus, a variable x_j with large variance will get a larger weight v_j , and hence give a larger influence on the latent variable u than a variable x_j with a small variance.

When no prior information about the relative importance of the independent variables is available, centering the variables x_j to mean zero and scaling to unit length is probably the best alternative. As noted by Draper and van Nostrand (1979): "This at least forces everyone to do the same calculations in circumstances where compelling prior information is lacking."

6. Cross validation and jackknife. Cross validation (Stone (1974) and Geisser (1974)) is a technique which is very useful in estimating the optimal complexity of a model for a given data set. The data set is divided into a number of groups. The model, with a given complexity, is fitted to the data set reduced by one of the groups. Predictions are calculated by the fitted model for the deleted data and the sum of squares of predicted minus observed values for the deleted data is formed. Then, in a second round, the same procedure is repeated but with the second group held out. Then a third round is performed, etc., until each data point has been held out once and only once. The total sum of squares of predictions minus observations then contains one term from each point. This sum, abbreviated PRESS, is a measure of the predictive power of the model with the given complexity for the given data set.

Cross validation has attractive theoretical properties (Wahba (1977)). Golub, Heath and Wahba (1979) use it to estimate the optimal ridge factor in ridge regression. It has been used as a criterion to select variables in multiple regression (Allen, 1971), for selecting the smoothing factor in spline fitting (Wahba and Wold (1974), Craven and Wahba (1979)), to select the best number of components in principal components analysis (S. Wold, 1978), and for hypothesis testing in PLS modelling (H. Wold (1982)). In the PLS estimation discussed here, we wish to estimate the optimal rank of the estimate, i.e. when to stop the algorithm. We divide the cases into G groups. With one of these groups deleted one still gets estimates for v_{s+1} in step (2.1) of algorithm LSQR, which allows the "latent" variable Xv_{s+1} in step (2.2) to be estimated for all n data points, including those deleted. The residual p_{s+1} is then estimated based on the retained points, and prediction errors computed for those deleted. The PRESS is then calculated as the sum of squares of the predicted residuals for the deleted points. A second part of the data set is then held out, squared prediction errors are added to the PRESS and so on.

Note that in algorithm LSQR we have scaled p_s to unit length. When comparing prediction errors in different steps, we use the unscaled residuals

$$p_{s\text{UNSC}} = (\rho_1 \rho_2 \cdots \rho_s) / (\theta_1 \cdots \theta_{s-1}) p_s.$$

An alternative way of performing cross validation, which is to be used when the

successive PRESS'es are to be compared with ridge-PRESS and ordinary least squares PRESS, is to make the total calculation up to the last significant PLS rank separately for each subgroup deletion.

In large problems, repeated calculations with deletion of one group after another may be too time-consuming. In such cases, however, there are usually so many data points that one with little loss can make a small part, say 5 or 10%, of the data a "test set." The estimation is then based on the remaining 90 or 95% of the data and the validation is made by letting the model predict the values in the test set. Thus the test set never enters the estimation; it is used only to estimate the "optimal rank" of the model.

The jackknife method (see Duncan (1978) and Miller (1974) for reviews) is closely related to cross validation in that it makes use of the data several times, each time with a subset of the data deleted from the calculation. The scope of the jackknife is to use the variation in the resulting parameter estimates $\hat{\beta}$ to calculate standard errors of these estimates. It is a "soft," data-oriented approach, in contrast to methods based on "hard" models such as maximum-likelihood methods (see H. Wold (1982) for a discussion). Denoting the value of $\hat{\beta}_j$ obtained when subgroup i ($i = 1, 2, \dots, I$) is deleted by $\hat{\beta}_{ji}$, one first forms the pseudo-values P_{ji} as:

$$P_{ji} = I \cdot \hat{\beta}_j - (I-1)\hat{\beta}_{ji} \quad (i = 1, 2, \dots, I).$$

These values have the averages

$$\bar{P}_j = \frac{1}{I} \sum_i P_{ji}$$

and the estimated standard errors:

$$s_j = \left[\frac{1}{I(I-1)} \sum (P_{ji} - \bar{P}_j)^2 \right]^{1/2}.$$

7. A chemical example. Dunn, Wold and Martin (1978) tried to relate the biological activity, y (the stimulation of the β -receptor measured by Lefkowitz et al., 1976), of $n = 16$ similar chemical compounds to a set of 8 variables x_j . These variables describe various morphological and physicochemical properties of these compounds, such as electronic and steric properties of the substituent in a certain position and the lipophilic character and receptor-binding strength of the whole molecule. The data are shown in Table 1.

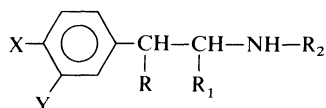
In their analysis, Dunn, Wold and Martin (1978) use principal components regression, contracting the matrix X to its first two singular vectors.

Here we reanalyze the same data using partial least squares, PLS. For comparison we have used ordinary least squares (OLS), principal component regression (PCR), James Stein shrunk estimate (JS), ridge regression (RR), and finally total least squares (TLS).

To evaluate the goodness of fit of the various models, we use the predictive sums of squares (PRESS) as described above in § 6. The data were divided into 3 groups (no. 1 = points 1, 4, 7, \dots , no. 2 = points 2, 5, 8, \dots , no. 3 = points 3, 6, 9, \dots). Thus, the jackknife standard errors were also obtained for the PLS estimates.

Figure 1 shows the variation of the ridge estimates, PRESS and the residual sum of squares with the ridge parameter k . PRESS has a minimum around $k = 2.0$, which is larger than the maximal recommended value of 1.0 (Hoerl and Kennard (1970)).

TABLE I
Observed β -receptor agonist activity (y) for 15 compounds of the formulae



The structural descriptors x_1 to x_8 are, in order: x_1 = log equilibrium constant for the binding of the molecule to a receptor model, x_2 = log acidity constant, x_3 = lipophilicity (tendency to prefer fatty tissues compared to aqueous phase, e.g. blood) of the phenyl part (X, Y—C₆H₄ in the formulae above) as measured on the model system octanol–water (see Hansch et al. (1973)), x_4 and x_5 are lipophilicities of groups R₁ and R₂, respectively, defined in the same way as x_3 , x_6 and x_7 are Taft's σ^* and E_s of the group R₂ which measure the electronic and steric properties of the group (see Hansch et al. (1973)) and x_8 is an indicator variable equal to 1 when R=OH and 0 when R=H.

| t | y | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 |
|-----|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 4.39 | 4.55 | 8.93 | 1.14 | 0.70 | 0.19 | 0.49 | 1.24 | 1 |
| 2 | 4.42 | 4.74 | 8.93 | 1.14 | 1.23 | 0.19 | 0.49 | 1.24 | 1 |
| 3 | 5.00 | 5.07 | 9.29 | 1.14 | 0.19 | 0.70 | 0.00 | 0.00 | 1 |
| 4 | 5.85 | 5.77 | 9.90 | 1.14 | 0.19 | 1.64 | -0.10 | -0.47 | 1 |
| 5 | 4.35 | 4.62 | 9.90 | 1.14 | 1.23 | 1.64 | -0.10 | -0.47 | 1 |
| 6 | 4.51 | 4.41 | 9.93 | 1.14 | 1.23 | 2.35 | -0.20 | -0.51 | 1 |
| 7 | 6.33 | 6.17 | 9.19 | 1.14 | 0.19 | 2.83 | -0.13 | -0.93 | 1 |
| 8 | 6.37 | 6.17 | 9.19 | 1.14 | 0.19 | 2.56 | -0.13 | -0.93 | 1 |
| 9 | 4.68 | 4.33 | 10.03 | 1.14 | 0.19 | 2.42 | -0.08 | -0.38 | 0 |
| 10 | 5.04 | 4.62 | 10.29 | 1.14 | 0.19 | 3.36 | -0.13 | -0.93 | 0 |
| 11 | 7.10 | 7.22 | 9.29 | 1.14 | 0.19 | 2.43 | -0.30 | -1.60 | 1 |
| 12 | 5.04 | 4.64 | 10.22 | 1.14 | 0.19 | 2.95 | -0.08 | -0.38 | 0 |
| 13 | 6.00 | 5.62 | 9.94 | -0.07 | 0.19 | 1.64 | -0.19 | -0.47 | 1 |
| 14 | 5.48 | 6.19 | 9.77 | -0.07 | 0.19 | 1.64 | -0.19 | -0.47 | 1 |
| 15 | 7.10 | 7.85 | 9.29 | -0.07 | 0.19 | 3.80 | -0.30 | -1.60 | 1 |

The OLS PRESS is 3.16 which is considerably higher than the best ridge value of 2.11. The PLS PRESS is 2.06, very close to the ridge optimum. The PCR value of 2.56 (with two components) is somewhat larger. The James–Stein estimates give a minimal PRESS = 2.57 for the shrinking factor of 0.82. The PCR PRESS for three components equals 2.12. TLS (minimum norm) gave PRESS = 2.02 for rank = 3, and behaved very similarly to PCR for low ranks. When full rank is approached, it behaved differently. Note that in order to get a minimum norm TLS solution, the complete singular value decomposition is needed (see Golub and Van Loan (1980, formula (2.7))).

We also tested cross validation with 5 and 15 groups, i.e. only one point deleted. The results are not significantly different; in some cases we got a sharper minimum for PRESS with 5 groups. Generally, few groups are expected to give a conservative estimate of complexity, since the risk of overfitting is smaller.

Table 2 shows the estimated values of the regression coefficients $\hat{\beta}_j$ for the various methods. It is seen that the Ridge, PLS and PCR values are shrunk compared with the OLS values, the PLS and PCR values being the most shrunk. The ridge, PCR (2) and the PLS (2) agree within about 2 standard errors of the latter.

8. Discussion. The PLS method gives a solution to the multiple regression problem which is stabilized in comparison with the OLS solution and which has, at least in the examples investigated, a comparable prediction error to ridge regression. The very simple computations involved makes the method suitable for large problems and

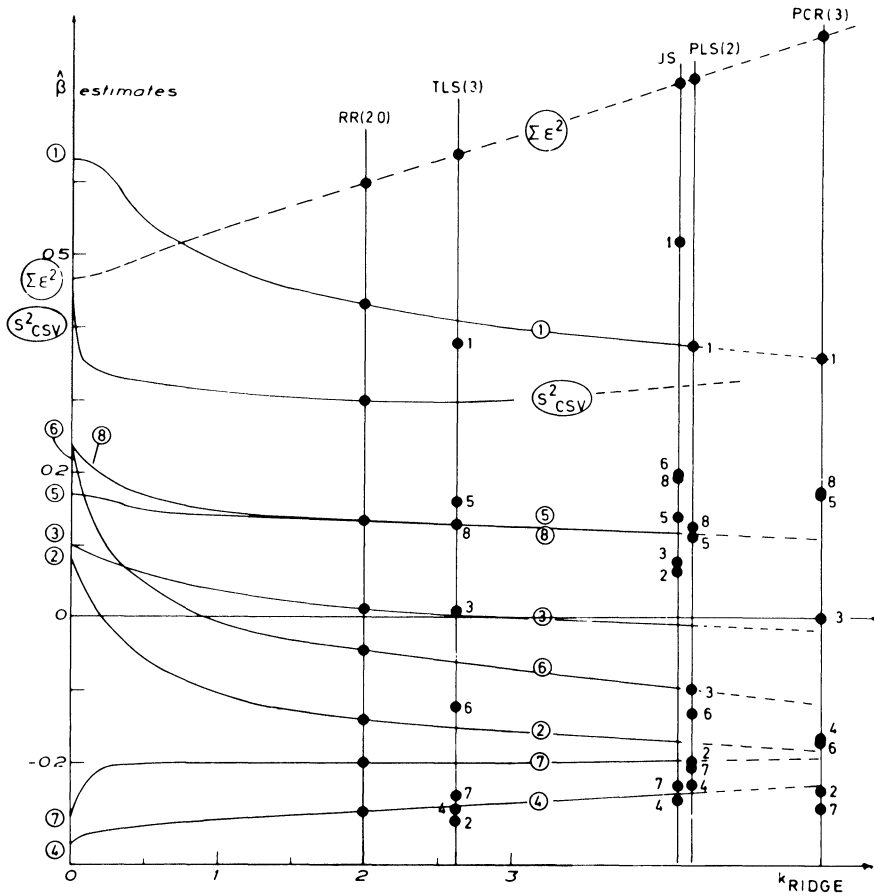


FIG. 1. Estimated values for regression coefficients 1 through 8 as functions of the ridge parameter k . Also plotted are the estimates corresponding to the TLS, JS, PLS and PCR solutions (see Table 2). The sum of squared residuals and the cross validation variance are plotted in relative scales.

TABLE 2

Resulting estimates of the regression coefficients β_i (data in Table 1*) for the different methods of estimation. SS is the residual sum of squares obtained when the model is fitted to all data. PRESS is the corresponding predictive sum of squares obtained by cross validation with the data divided into three groups. OLS is ordinary least squares, JS is James–Stein estimate with shrinking factor 0.82, RR ridge regression with $k = 2.0$, PCR principal components regression with 2 and 3 components, TLS is total least squares and PLS is partial least squares rank 1 and 2, respectively. The PLS(2) standard errors (s.err) were calculated by the jackknife method.

| Var. no | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | SS | PRESS |
|----------|------|-------|-------|-------|------|-------|-------|------|------|-------|
| OLS | .636 | .080 | .095 | -.308 | .169 | .241 | -.278 | .238 | 0.61 | 3.156 |
| JS (.82) | .522 | .066 | .078 | -.252 | .138 | .197 | -.228 | .195 | 1.04 | 2.569 |
| RR (2.0) | .432 | -.143 | .011 | -.267 | .133 | -.049 | -.200 | .133 | 0.82 | 2.114 |
| PCR (2) | .325 | -.121 | -.210 | -.144 | .111 | -.184 | -.210 | .194 | 1.99 | 2.557 |
| PCR (3) | .354 | -.235 | .018 | -.162 | .168 | -.164 | -.262 | .172 | 1.15 | 2.117 |
| TLS (3) | .379 | -.274 | .030 | -.260 | .161 | -.125 | -.244 | .132 | 0.88 | 2.015 |
| PLS (1) | .284 | -.069 | -.122 | -.192 | .161 | -.184 | -.223 | .085 | 2.15 | 3.030 |
| PLS (2) | .372 | -.205 | -.090 | -.226 | .119 | -.128 | -.211 | .166 | 1.06 | 2.062 |
| s.err | .058 | .081 | .103 | .046 | .009 | .047 | .016 | .045 | | |

* All variables including y scaled to mean zero and unit variance.

for the implementation on micro-computers and desk calculators. Cross validation provides a simple and straightforward stopping rule; it also makes it simple to compare the estimates of different methods and to calculate jackknife standard errors.

The situation treated in the present article is a simple special case of the general PLS modelling and estimation. The favorable predictive properties of the PLS estimates found in this case is encouraging for the utility of more extended PLS models (cf. H. Wold, 1982).

We have been using the principal components of X as a means of discriminating between different classes of "objects" (cases, compounds), each object being characterized by values of the variables x_j (Albano et al. (1978), Dunn et al. (1978), (1979), S. Wold et al. (1976), (1977), (1978)). The extra information provided by the variable y in the present case might make the corresponding vectors b_s superior to the ordinary principal components in the context of discriminant analysis and pattern recognition, a subject which we are currently investigating.

In a forthcoming report we will also investigate statistical and numerical aspects on the PLS method, and show how the 8 theorems given by Marquardt (1970) for ridge and principal component regression translate into this situation. We will also investigate computational aspects, especially for large problems. For large sparse problems, e.g., the geodesy problem, a reformulation as a multiple block PLS problem might prove advantageous.

Acknowledgments. The present project is supported by grants from the Swedish Natural Science Research Council (NFR), the Swedish Board of Research Councils (FRN) and the Stiftung Volkswagenwerk. Helpful discussions with Carl-Ivar Brändén and Åke Björck are greatly appreciated.

REFERENCES

- [1] C. ALBANO, W. J. DUNN, U. EDLUND, E. JOHANSSON, B. NORDEN, M. SJÖSTRÖM AND S. WOLD, (1978), *Four levels of pattern recognition*, Anal. Chim. Acta. Comput. Techn. Optim., 2, pp. 429–443.
- [2] D. M. ALLEN, (1971), *Mean square error of prediction as criterion for selecting variables*, Technometrics, 13, pp. 469–475.
- [3] Å. BJÖRCK AND L. ELDEN, (1979), *Methods in numerical algebra for ill-posed problems*, Tech. Rep. MAT-R-33-1979, Linköping University, Sweden.
- [4] G. E. P. BOX AND N. R. DRAPER, (1971), *Factorial designs, the $|X'X|$ criterion and some related matters*, Technometrics, 13, pp. 731–742.
- [5] P. CRAVEN AND G. WAHBA, (1979), *Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross validation*, Numer. Math., 31, pp. 377–403.
- [6] N. R. DRAPER AND H. SMITH, (1981), *Applied Regression Analysis*, 2nd ed., John Wiley, New York.
- [7] N. R. DRAPER AND R. C. VAN NOSTRAND, (1979), *Ridge regression and James–Stein estimation. Review and comments.*, Technometrics, 21, pp. 451–466.
- [8] G. T. DUNCAN, (1978), *An empirical study of jackknife constructed confidence regions in non-linear regression*, Technometrics, 20, pp. 123–129.
- [9] W. J. DUNN III, S. WOLD AND Y. MARTIN, (1978), *Structure activity study of beta adrenergic agents with the SIMCA method*, J. Med. Chem., 21, pp. 922–930.
- [10] W. J. DUNN III AND S. WOLD, (1980), *Relationships between chemical structure and biological activity modelled by SIMCA pattern recognition*, Bioorg. Chem., 9, pp. 505–523.
- [11] S. GEISSER, (1974), *A predictive approach to the random effect model*, Biometrika, 61, pp. 101–107.
- [12] G. H. GOLUB AND W. KAHAN, (1965), *Calculating the singular values and pseudo-inverses of a matrix*, SIAM J. Numer. Anal., 2, pp. 205–224.
- [13] G. H. GOLUB, M. HEATH AND H. WAHBA, (1979), *Generalized cross validation as a method for choosing a good ridge parameter*, Technometrics, 21, pp. 215–224.
- [14] G. H. GOLUB AND C. F. VAN LOAN, (1980), *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17, pp. 883–893.

- [15] J. W. GORMAN AND R. J. TOMAN, (1966), *Selection of variables for fitting equations to data*, Technometrics, 8, pp. 27–51.
- [16] C. HANSCH, S. UNGER, K. H. KIM, D. NIKAITANI AND E. J. LIEN, (1973), *Aromatic substituent constants for structure-activity correlations*, J. Med. Chem., 16, pp. 120–126.
- [17] D. M. HAWKINS, (1975), *Relations between ridge regression and eigenanalysis of the augmented correlation matrix*, Technometrics, 17, pp. 477–480.
- [18] M. R. HESTENES AND E. STIEFEL, (1952), *Methods of conjugate gradients for solving linear systems*, J. Res. NBS, 49, pp. 409–436.
- [19] R. R. HOCKING, (1976), *The analysis and selection of variables in linear regression*, Biometrics, 32, pp. 1–49.
- [20] A. E. HOERL AND R. W. KENNARD, (1970), *Ridge regression: biased estimates for non-orthogonal problems*, Technometrics, 12, pp. 55–67.
- [21] G. B. KOLATA, (1978), *Geodesy: Dealing with an enormous computer task*, Science 200, April 28, pp. 421–422.
- [22] J. H. KONNERT, (1976), *A restrained parameter structure sector least squares procedure for large asymmetric units*, Acta Cryst., A32, pp. 614–617.
- [23] D. W. MARQUARDT, (1970), *Generalized inverses, ridge regression, biased linear and non-linear estimation*, Technometrics, 12, pp. 591–612.
- [24] L. S. MAYER AND T. A. WILLKE, (1973), *On biased estimation in linear models*, Technometrics, 15, pp. 497–508.
- [25] R. G. MILLER, (1974), *The jackknife, a review*, Biometrika, 61, pp. 1–15.
- [26] C. MUKHERJEE, M. C. CARON, D. MULLIKEN AND R. J. LEFKOWITZ, (1976), *Structure-activity relationships of adenyl cyclase coupled beta-adrenergic receptors: determination by direct binding studies*, Molecular Pharmacology, 12, pp. 16–31.
- [27] C. C. PAIGE AND M. A. SAUNDERS, (1982), *A bidiagonalization algorithm for sparse linear equations and least squares problems*, ACM Trans. Math. Software, 8, pp. 43–71.
- [28] G. SMITH AND F. CAMPBELL, (1980), *A critique of some ridge regression methods (with comments)*, J. Amer. Statist. Assoc., 75, pp. 74–103.
- [29] M. STONE, (1974), *Cross-validatory choice and assessment of statistical predictions*, J. Roy. Statist. Soc., 36, pp. 111–133.
- [30] J. M. VARAH, (1979), *A practical examination of some numerical methods for linear discrete ill-posed problems*, SIAM Rev., 21, pp. 100–111.
- [31] G. WAHBA, (1977), *Practical approximate solution to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., 14, pp. 651–657.
- [32] G. WAHBA AND S. WOLD, (1975), *A completely automatic French curve: Fitting spline functions by cross-validation*, Comm. Statist., 4, pp. 1–17.
- [33] H. WOLD, (1975), *Soft modelling by latent variables; the non-linear iterative partial least squares approach*, in Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett, J. Gani, ed., Academic Press, London.
- [34] ———, (1977), *On the transition from pattern recognition to model building*, in Mathematical Economics and Game Theory, Essays in Honour of Oscar Morgenstern, R. Hern and O. Moeschlin, eds., Springer-Verlag, Berlin.
- [35] ———, (1982), *Soft modeling, The basic design and some extensions*, In Systems Under Indirect Observation, I-II, K. G. Jöreskog and H. Wold, eds., North-Holland, Amsterdam.
- [36] S. WOLD, (1976), *Pattern recognition by means of disjoint principal components models*, Pattern Recogn., 8, pp. 127–139.
- [37] ———, (1978), *Cross validatory estimation of the number of components in factor and principal components models*, Technometrics, 20, pp. 397–405.
- [38] S. WOLD AND M. SJÖSTRÖM, (1977), *SIMCA, a method for analyzing chemical data in terms of similarity and analogy*, in Chemometrics, Theory and Practice, B. R. Kowalski, ed., ACS Publ. Ser. 52, American Chemical Society, pp. 243–282.