



ELSEVIER

Linear Algebra and its Applications 321 (2000) 295–305

LINEAR ALGEBRA  
AND ITS  
APPLICATIONS

www.elsevier.com/locate/laa

# Ill-conditioning and multicollinearity

Fikri Öztürk<sup>a,\*</sup>, Fikri Akdeniz<sup>b</sup>

<sup>a</sup>Department of Statistics, Ankara University, Ankara, Turkey

<sup>b</sup>Department of Mathematics, Çukurova University, Adana, Turkey

Received 12 June 1997; accepted 20 April 2000

Submitted by H.J. Werner

## Abstract

It is well known that unstability of solutions to small changes in inputs causes many problems in numerical computations. Existence, uniqueness and stability of solutions are important features of mathematical problems. Problems that fail to satisfy these conditions are called ill-posed. The purpose of this study is to remind briefly some methods of solution to ill-posed problems and to see the impacts or connections of these techniques to some statistical methods. © 2000 Elsevier Science Inc. All rights reserved.

*AMS classification:* 65F35; 62J07

*Keywords:* Ill-posed problem; Ill-conditioning; Multicollinearity

## 1. Introduction

In general, a system (or problem) has an input (initial value, data)  $u$  and an output (solution)  $r$ , where  $r = R(u)$  and  $R$  is a method of solution.

Let  $u$  be in the metric space  $U$  with a metric  $d_U$  and  $r = R(u)$  be in the metric space  $F$  with a metric  $d_F$ . Then the problem is called stable in the pair of spaces  $(F, U)$  if for all  $\epsilon > 0$  there exists  $\delta(\epsilon) > 0$  such that

$$d_U(u_1, u_2) \leq \delta(\epsilon) \Rightarrow d_F(r_1, r_2) \leq \epsilon,$$

where  $u_1, u_2 \in U$  and  $r_1 = R(u_1)$ ,  $r_2 = R(u_2)$ .

\* Corresponding author.

*E-mail addresses:* ozturk@turkuaz.kku.edu.tr (F. Öztürk), akdeniz@mail.cu.edu.tr (F. Akdeniz).

The problem of determining the solution  $r$  in the space  $F$  from the initial data  $u$  in the space  $U$  is said to be well posed on the pair of metric spaces  $(F, U)$  if the following three conditions are satisfied:

- (i) for every element  $u \in U$ , there exists a solution  $r$  in  $F$ ;
- (ii) the solution is unique;
- (iii) the problem is stable on  $(F, U)$ .

Problems that fail to satisfy these conditions are called ill-posed. Problems that satisfy conditions (i) and (ii) but not (iii) are called ill-conditioned, that is, problems which are unstable are called ill-conditioned [1,2,20].

It should be pointed out that the definition of an ill-posed problem is related to the pair of metric spaces  $(F, U)$ . The same problem may be well posed in other metrics.

As a simple example let us consider the following two systems of equations:

$$\begin{cases} \frac{1}{6}x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \frac{1}{3}y + \frac{2}{3}z = \frac{4}{3}, \\ 0.4999x + 0.4999y + z = 1.999 \end{cases}$$

with solutions

$$x = 1, \quad y = 1, \quad z = 1,$$

and

$$\begin{cases} \frac{1}{6}x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \frac{1}{3}y + \frac{2}{3}z = \frac{4}{3}, \\ 0.4999x + 0.4999y + z = 1.998 \end{cases}$$

with solutions

$$x = 10, \quad y = 10, \quad z = -8.$$

The only difference between these systems is the value of the third component in the right-hand side vectors. If we take the right-hand side vectors as inputs  $u_1, u_2 \in U = R^{3 \times 1}$ , the inverse of the coefficient matrix as  $R$  and the solutions as  $r_1, r_2 \in F = R^{3 \times 1}$ , then according to the Euclidean metric

$$d_U(u_1, u_2) = \sqrt{(1-1)^2 + \left(\frac{4}{3} - \frac{4}{3}\right)^2 + (1.9998 - 1.998)^2} = 0.0018$$

and

$$d_F(r_1, r_2) = \sqrt{(1-10)^2 + (1-10)^2 + (1-(-8))^2} = 9\sqrt{3}.$$

Small changes in inputs result large changes in outputs.

In order to see that the problem is well posed, we need to check two things. First one is the existence and uniqueness and the second one is the stability of the solution.

For a real linear equation system it is easy to check the existence and uniqueness of the solution. But it is more difficult to check the stability. Let us focus on the latter problem. For a nonsingular matrix  $A \in R^{n \times n}$  and  $b \in R^{n \times 1}$  the solution of the system  $Ax = b$  exists and is unique. Suppose that  $\Delta A$  and  $\delta b$  are, respectively, the perturbations of  $A$  and  $b$  in the linear system  $Ax = b$ . If  $A$  is replaced by a nonsingular matrix  $A + \Delta A$  and  $b$  is replaced by  $b + \delta b$ , then the error (change) in the solution can be bounded as

$$\|\Delta x\| \leq \frac{\|A^{-1}\|(\|\Delta b\| + \|\Delta A\| \|x\|)}{1 - \|A^{-1}\| \|\Delta A\|}$$

for  $\|A^{-1}\| \|\Delta A\| < 1$  (see e.g. [2, p. 404] and [8, 10]). When there is no change in  $A$ , that is, when  $\Delta A = 0$ , then

$$\|\Delta x\| \leq \|A^{-1}\| \|\Delta b\|.$$

The relationship between the relative error  $\|\Delta x\|/\|x\|$  and the relative change on the right-hand side vector  $\|\Delta b\|/\|b\|$  is

$$\frac{\|\Delta x\|}{\|x\|} = k(A) \frac{\|\Delta b\|}{\|b\|},$$

where  $k(A)$  is the condition number of  $A$  and is defined by

$$k(A) = \sup_b \left[ \sup_{\Delta b} \left( \frac{\|\Delta x\|}{\|x\|} / \frac{\|\Delta b\|}{\|b\|} \right) \right] = \|A^{-1}\| \|A\| \geq \frac{\max |\lambda_A|}{\min |\lambda_A|},$$

where  $|\lambda_A|$  is the absolute value of an eigenvalue  $\lambda_A$  of  $A$ .

When the norm is the Euclidean norm and  $A$  is symmetric, then

$$k(A) = \frac{\max |\lambda_A|}{\min |\lambda_A|}.$$

If the condition number  $k(A)$  of the coefficient matrix in the linear equation system  $Ax = b$  is large, then the problem with input  $b$  and output  $x = A^{-1}b$  is unstable, that is, the problem (inverse problem) is ill-conditioned. The matrix  $A$  itself is called ill-conditioned.

## 2. Some solution methods for ill-posed inverse problems

In general, an inverse problem contains a known operator  $A$ , a function (a vector)  $z$  which is a characterization of the phenomenon that we are going to model and input  $u$  obtained from the measurements, such that  $Az = u$ . The purpose of the inverse problem is to solve for  $z$ . Since the input comes from measurements, we have some  $\tilde{u}$  for  $u$ . If  $\tilde{u}$  does not belong to the range of  $A$ , then there is a problem of existence of a solution. Even if  $\tilde{u}$  is in the range of  $A$  and the solution is unique it may not be stable, that is, the inverse problem may be ill-conditioned. So, measurement errors will cause some problems. Also, when the inverse problem is ill-conditioned and  $\tilde{u}$

is an observation of a random function (or vector), then the solution (for example, the estimate) will be unreliable.

Now we will remind some solution methods for ill-posed inverse problems from the book [20] by Tikhonov and Arsenin. It will be supposed that the operator  $A$  has a closed range.

### 2.1. Selection method

For  $z$  in  $F$  with a metric  $d_F$ ,  $u$  in  $U$  with a metric  $d_U$  and an operator  $A : F \rightarrow U$  let us consider the inverse problem  $Az = \tilde{u}$ . Then the selection method for ill-posed inverse problems consists of a selection of a subset  $V$  ( $V \subset F$ ) and search for a solution  $z_0$  ( $z_0 \in V$ ) such that

$$\inf_{z \in V} d_U(Az, \tilde{u}) = d_U(Az_0, \tilde{u}).$$

When  $F, U$  are linear spaces,  $A : F \rightarrow U$  is a linear operator and  $U$  is a Hilbert space, for a subspace  $V$  ( $V \subset F$ ),

$$Az_0 = P_{A(V)}u, \quad (1)$$

where  $P_{A(V)}$  is the orthogonal projection onto  $A(V)$ .

In this method, the question becomes the selection of the subspace  $V$  in order to overcome the ill-posedness, that is, the existence of a solution and (or) its instability.

### 2.2. Replacement method

In this method, the operator  $A$  in the equation  $Az = \tilde{u}$  is replaced by a regular operator  $A + \delta I$  ( $\delta \in R, \delta > 0$  and  $I$  is the identity operator) and

$$z_\delta = (A + \delta I)^{-1}\tilde{u} \quad (2)$$

is taken to be a solution. The question is how to choose the regularization parameter  $\delta$ .

### 2.3. Regularization method

In this method, in order to stabilize the solution a continuous functional

$$\Psi : F_1 \subset F \rightarrow R^+ \cup \{0\}$$

is defined and the functional

$$M_{\Psi, \alpha}(z, \tilde{u}) = d_U^2(Az, \tilde{u}) + \alpha \Psi(z)$$

is constructed, for a chosen value  $\alpha$  ( $\alpha > 0$ ). The functional  $M_{\Psi, \alpha}$  is called smoothing functional for the inverse problem  $Az = \tilde{u}$ . The element  $z_\alpha$  in  $F_1$  which minimizes the smoothing functional is called regularized solution.

$$\inf_{z \in F_1} M_{\Psi, \alpha}(z, \tilde{u}) = M_{\Psi, \alpha}(z_\alpha, \tilde{u}).$$

The regularized solution  $z_\alpha$  is a function of  $\tilde{u}$  and  $\alpha$  and it also depends on  $\Psi$ . So, the question is how to choose  $\Psi$  and  $\alpha$ .

When  $F, U$  are Hilbert spaces,  $A : F \rightarrow U$  is a linear operator and  $\Psi$  is a functional defined by

$$\Psi : F \rightarrow R^+ \cup \{0\},$$

$$z \rightarrow \Psi(z) = \|z\|^2,$$

then the smoothing functional is

$$M_\alpha(z, \tilde{u}) = \|Az - \tilde{u}\|^2 + \alpha\|z\|^2 = (Az - \tilde{u})^*(Az - \tilde{u}) + \alpha z^*z, \quad (3)$$

where  $*$  denotes the conjugate transpose. In order to minimize  $M_\alpha(z, \tilde{u})$ , the derivative with respect to  $z$  is set to zero and the following equation is obtained:

$$(A^*A + \alpha I)z = A^*\tilde{u}.$$

The solution of this equation is the regularized solution

$$z_\alpha = (A^*A + \alpha I)^{-1} A^*\tilde{u} = \sum_{n=1}^{\infty} \frac{c_n}{\lambda_n + \alpha} \varphi_n, \quad (4)$$

where  $\lambda$ 's and  $\varphi$ 's are eigenvalues and eigenvectors of  $A^*A$ , respectively, and the coefficients  $c_n$  are such that  $A^*\tilde{u} = \sum_{n=1}^{\infty} c_n \varphi_n$ .

#### 2.4. Iterative method

Assume that the inverse problem  $Az = \tilde{u}$  has a unique solution. An iterative method is given by

$$z_{n+1} = z_n - hA^*(Az_n - \tilde{u}), \quad n = 0, 1, 2, \dots, \quad (5)$$

where  $z_0$  ( $z_0 \in F$ ) is an initial value and  $h$  is a number such that  $0 < h \|A^*A\| < 2$  [4].

In this method,  $h$  can be chosen differently for every iteration or  $h$  can be replaced by an operator  $H$ .

By applying the replacement method to preceding equation system the following solutions have been obtained:

*Eureka—the solver, version 1.0:*

$$\begin{cases} \frac{1}{6}x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \frac{1}{3}y + \frac{2}{3}z = 4/3, \\ 0.4999x + 0.4999y + z = 1.9998 \end{cases}$$

with solutions

$$x = -0.00080016003, \quad y = .00040008002, \quad z = 2.0000000;$$

*Eureka—the solver, version 1.0:*

$$\begin{cases} \left(\frac{1}{6} + 0.1\right)x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \left(\frac{1}{3} + 0.1\right)y + \frac{2}{3}z = \frac{4}{3}, \\ 0.4999x + 0.4999y + (1 + 0.1)z = 1.9998 \end{cases}$$

with solutions

$$x = -0.52670350, \quad y = 1.0534070, \quad z = 1.5786372;$$

*Eureka—the solver, version 1.0:*

$$\begin{cases} \left(\frac{1}{6} + 0.001\right)x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \left(\frac{1}{3} + 0.001\right)y + \frac{2}{3}z = 4/3, \\ 0.4999x + 0.4999y + (1 + 0.001)z = 1.9998 \end{cases}$$

with solutions

$$x = 0.81628466, \quad y = 0.81381356, \quad z = 1.187302;$$

*Eureka—the solver, version 1.0:*

$$\begin{cases} \left(\frac{1}{6} + 0.00000001\right)x + \frac{2}{6}y + \frac{3}{6}z = 1, \\ \frac{1}{3}x + \left(\frac{1}{3} + 0.00000001\right)y + \frac{2}{3}z = \frac{4}{3}, \\ 0.4999x + 0.4999y + (1 + 0.00000001)z = 1.9998 \end{cases}$$

with solutions

$$x = 0.99997502, \quad y = 0.99997499, \quad z = 1.0000250.$$

In the above computer output, it is seen that the solutions are dependent on the values of the regularization parameter  $\delta$  and they are getting better for small values of  $\delta$ .

### 3. The problem of multicollinearity

In statistics, our inputs (data) are observed values. Outputs are values of some statistics. When the outputs are sensitive (unstable) to small changes in data we can consider this sensitivity as an ill-conditioned problem. The variance of the statistic is a natural measure of this sensitivity.

In statistical inference, the statistics (estimators) are obtained according to some optimization criteria or principles. The best statistics under some criteria may be sensitive (unstable) to data even though the criteria is the minimization of the variance. For example, in linear models, the Gauss–Markov estimator of any estimable

linear combination of the parameter vector has minimum variance among all linear unbiased estimators, but because of a bad design (multicollinearity) the variance may be large and the estimated value may be far distant from the true value.

In this section, we will give a short summary of the problem of ill-conditioning in linear models, that is, the problem of multicollinearity. We especially will emphasize that the origins of methods to handle the multicollinearity are in the solution techniques of ill-posed problems.

The multicollinearity is defined as the existence of nearly linear dependency among column vectors of the design matrix  $X = [\underline{X}_1, \underline{X}_2, \dots, \underline{X}_p]$  in the linear model  $\underline{Y} = X\underline{\beta} + \varepsilon$ . The existence of multicollinearity may result in wide confidence intervals for individual parameters (unstable estimates), may give estimates with wrong signs and may affect our decision in a hypothesis testing. Severe multicollinearity may make the estimates so unstable that they are practically useless.

Multicollinearity is synonymous with ill-conditioning. So, the problem of multicollinearity is another version of the problem of ill-conditioning in the solution of the following normal equations:

$$X'X\hat{\underline{\beta}} = X'\underline{Y}.$$

Therefore, in this inverse problem the condition number  $k(X'X)$  is a measure of existence of multicollinearity. Even though the condition number gives some information about the existence of multicollinearity, it does not explain the structure of the linear dependency among the column vectors  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_p$ . The best way of explaining the existence and structure of multicollinearity is to look at the eigenvalues and eigenvectors of the matrix  $X'X$  [9,11,13,18].

Let the column vectors  $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_p$  be standardized such that  $X'X$  is in the form of a correlation matrix. Let the spectral decomposition of  $X'X$  be

$$X'X = V \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} V' = V D V' = \sum_{i=1}^p \lambda_i \underline{v}_i \underline{v}_i',$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ . If the condition number  $k(X'X) = \lambda_1/\lambda_p$  is large, then the multicollinearity exists and its structure is explained by eigenvector  $\underline{v}_p$  as

$$[\underline{X}_1, \underline{X}_2, \dots, \underline{X}_p] \underline{v}_p \approx 0.$$

In the last one to two decades, there are a number of papers on the cause of multicollinearity. It is found that some observations may be influential on the value of condition number. The omission of such observations may create or remove the problem of multicollinearity [3,5–7,23].

Hundreds of studies have been done on the problem of multicollinearity. Even though many solution methods have been proposed, the problem has not been solved completely. Among the methods, the most popular one is the ridge method proposed by Hoerl and Kennard [12]. The ordinary ridge estimator is

$$\hat{\underline{\beta}}_k = (X'X + kI)^{-1} X'Y \quad (6)$$

obtained by adding a small positive number  $k$  to the diagonal elements of the normal equations. Notice that this estimator has the same expression as the solution in the replacement method (2) and regularization method (4).

Ridge method has been a popular estimation method for over 25 years. It can yield more reliable point estimates of the coefficients than the ordinary least-squares estimates by reducing the standard errors of the estimates at the expense of introducing some bias. However, there are some problems with the choice of the biasing parameter  $k$ .

The addition of a small number to the diagonal elements of a matrix was also familiar to Marquardt [14], who used it in his algorithms for nonlinear optimization. Marquardt defined a class of generalized inverse estimators and discussed their properties in the frame of biased estimators including the ordinary ridge estimator [15].

Let us now consider the normal equations  $X'X\hat{\underline{\beta}} = X'Y$  with ill-conditioned matrix  $X'X$  and let us try to solve the normal equations by the selection method. For a subspace  $K$  ( $K \subset R^p$ ) we look for a solution  $\hat{\underline{\beta}}_0 \in K$  such that

$$\min_{\hat{\underline{\beta}} \in K} d_E(X'X\hat{\underline{\beta}}, X'Y) = d_E(X'X\hat{\underline{\beta}}_0, X'Y),$$

where  $d_E$  is the Euclidean metric on  $R^p$ . From (1)

$$X'X\hat{\underline{\beta}}_0 = P_{X'X(K)}(X'Y). \quad (7)$$

When  $K = \text{span}\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_m\}$ , that is,  $K$  is spanned by eigenvectors corresponding to  $m$  largest eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_m$  with  $m < p$ , then  $X'X(K) = K$  and

$$P_{X'X(K)} = P_K = W(WW')^{-1}W' = WW',$$

where  $W = [\underline{v}_1, \underline{v}_2, \dots, \underline{v}_m]$ . Moreover, from (7)

$$\begin{aligned} \hat{\underline{\beta}}_0 &= (X'X)^{-1} WW'X'Y = VD^{-1}V'WW'X'Y \\ &= VD^{-1} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} W'X'Y = V \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} V'X'Y \end{aligned}$$

This estimator  $\hat{\underline{\beta}}_0$  obtained by the selection method is the principle component estimator of  $\underline{\beta}$ . In the selection method, the trouble was selecting the subspace for the solutions. Here the trouble is in choosing the integer  $m$ , the number of principal



components. There are numerous suggestions in the literature concerning appropriate criteria for retention or deletion of components.

Estimators based on the iterative method (5) and their statistical properties are also investigated. An estimation procedure based on the iterative method can be constructed as follows. Let  $\hat{\beta}_{(0)}$  be a fixed point in the parameter space and

$$\hat{\beta}_{(n)} = (I - hX'X)\hat{\beta}_{(n-1)} + hX'Y, \quad n = 1, 2, \dots,$$

where  $h \in (0, 1/\lambda_1)$ . For  $\hat{\beta}_{(0)} = \underline{0}$  and suitable chosen values  $h, n$  the estimator  $\hat{\beta}_{(n)}$  can have smaller mean-squared error than the least-squares estimator [17]. An estimator similar to  $\hat{\beta}_{(n)}$  is proposed by Trenkler [21]. Trenkler's motivation is based on a serial expansion for the Moore–Penrose generalized inverse.

The estimators considered above are better than the ordinary least-squares estimator in the case of multicollinearity according to the mean-squared error criteria.

Nowadays the problem of multicollinearity is also a subject of research in generalized and multivariate linear models.

The ridge method, which corresponds to operator replacement method for ill-conditioned problems, is used extensively in linear models as a solution method to overcome multicollinearity. The ridge method is also used in statistical analysis techniques which involves matrix inverses. For example, the Mahalanobis distance

$$D(\underline{x} - \underline{\mu}) = (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$$

involves a matrix inversion, where  $\underline{x}$  is an observed vector,  $\underline{\mu}$  and  $\Sigma$  are the mean vector and covariance matrix of the distribution (population). Substituting  $\Sigma + kI$  for  $\Sigma$ ,

$$D_k(\underline{x} - \underline{\mu}) = (\underline{x} - \underline{\mu})' (\Sigma + kI)^{-1} (\underline{x} - \underline{\mu})$$

is called the ridge type Mahalanobis distance. In discriminant analysis applications the classification rule based on the Mahalanobis distance involves sample covariance matrix  $\hat{\Sigma}$ . By substituting  $\hat{\Sigma} + kI$  for  $\hat{\Sigma}$  a ridge type classification rule is obtained [19].

As another example let us consider the canonical correlation analysis:

$$\begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}_{p \times 1} \sim N \left( \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

$$V_1 = \sum_{i=1}^p \alpha_i X_{1i}, \quad V_2 = \sum_{j=1}^k \beta_j X_{2j}.$$

Canonical correlation analysis problem is to find the coefficients  $\alpha$ 's and  $\beta$ 's such that the correlation coefficient between the random variables  $V_1$  and  $V_2$  is maximized. Hottelling proved that the problem becomes the eigenvalue and eigenvector analysis of  $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ . In applications, the ridge type canonical correlation analysis [22] is based on  $(\hat{\Sigma}_{22} + k_2 I)^{-1} \hat{\Sigma}_{21} (\hat{\Sigma}_{11} + k_1 I)^{-1} \hat{\Sigma}_{12}$ .

In computations which frequently involve matrix inversion, as in control theory and nonlinear models, the ridge method can be used as a powerful tool against ill-conditioning.

#### **4. Conclusion**

A statistical analysis can be affected by ill-conditioning in three ways (stages). First, measurement errors can create some problems in the presence of ill-conditioning, since the outputs will be sensitive (unstable) to small changes in the inputs. At this stage the problem is the sensitivity of measurement devices and it looks more technological rather than statistical. Although statisticians ignore the measurement errors, sometimes they consider models with errors in variables. When errors are included in some variables, the statistical inference procedures will require some information about the distributional properties of these errors.

Second, inference procedures based on the optimization of some statistical criteria can give misleading results in the presence of ill-conditioning, caused by bad design or sampling. Because of randomness there always will exist a natural variability in the observations. Therefore, it is necessary to check the existence of ill-conditioning before performing the statistical analysis. That is, in some way, we have to check whether the values of statistics (outputs) concerned in an inference procedure are sensitive to small changes in the data (inputs).

Third, rounding errors (small changes in the inputs) can create problems in statistical computations when ill-conditioning exists. This computational aspect of the ill-conditioning is a purely numerical problem and is the most difficult one in numerical analysis. Perhaps, because of this, “Why has no one noticed” asked McCullough and Vinod [16]. They also added “It is understandable that economists have paid little attention to whether or not econometric software is accurate. Until recently, econometrics texts rarely discussed computational aspects of solving econometric problems. . . Many textbooks convey the impression all one has to do is use a computer to solve the problem, the implicit and unwarranted assumptions being that the computer’s solution is accurate and that one software package is as good as any other. . . pencil-and-paper arithmetic is not like computer arithmetic”.

To overcome the problems mentioned we can use some robust numerical algorithms and statistical procedures, make operations like in variable selection or outlier detection techniques or apply some “cure methods” like in ridge regression.

Finally, let us remind that the problem associated with ill-conditioning is a problem of severity rather than existency. The method of computation and the method of statistical inference greatly affect the results. In the last 40 years, some known numerical methods were adopted to obtain some statistical methods for inference in the presence of ill-conditioning. Some of the pioneers are Marquardt [15], Hoerl and Kennard [12], etc. Conversely, some results obtained by statisticians, for example, the research on leverage and influential observations, are helpful in creating

numerical algorithms and performing numerical computations when ill-conditioning exists. In the frame of software reliability statistical methods play a great role in the development studies of numerical methods which are robust against ill-conditioning.

## Acknowledgment

The authors would like to thank the associate editor and the referees for their helpful comments on this paper.

## References

- [1] K.E. Atkinson, *An Introduction to Numerical Analysis*, Wiley, New York, 1978.
- [2] N.S. Bakhvalov, *Numerical Methods*, Mir Publishers, Moscow, 1977.
- [3] D.A. Belsley, E. Kuh, R.E. Welsch, *Regression Diagnostics*, Wiley, New York, 1980.
- [4] M. Brill, E. Schock, Iterative solution of ill-posed problems – a survey, in: *Proceedings of the Fourth International Mathematical Geophysics Seminar*, Berlin, 1987.
- [5] S. Chatterjee, A.S. Hadi, Influential observation high leverage points and outliers in linear regression, *Statist. Sci.* 1 (1986) 379–416.
- [6] S. Chatterjee, A.S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York, 1988.
- [7] R.D. Cook, S. Weisberg, *Residuals and Influence in Regression*, Chapman & Hall, New York, 1982.
- [8] B.N. Datta, *Numerical Linear Algebra and Applications*, Brooks/Cole Publishing Company, 1994.
- [9] J. Fox, G. Monette, Generalized collinearity diagnostics, *J. Amer. Statist. Assoc.* 87 (1992) 178–183.
- [10] G.H. Golub, Ch.F. Van, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1991.
- [11] R.F. Gunst, Regression analysis with multicollinear predictor variables: definition detection and effects, *Comm. Statist. Theory Methods* 12 (1983) 2217–2260.
- [12] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation of nonorthogonal problems, *Technometrics* 12 (1970) 55–67.
- [13] E.R. Mansfield, B.P. Helms, Detecting multicollinearity, *Amer. Statist.* 36 (1982) 158–160.
- [14] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. App. Math.* 11 (1963) 431–441.
- [15] D.W. Marquardt, Generalized inverses ridge regression biased linear estimation and nonlinear estimation, *Technometrics* 12 (1970) 591–612.
- [16] B.D. McCullough, H.D. Vinod, The numerical reliability of econometric software, *J. Econom. Literature* XXXVII (1999) 633–665.
- [17] F. Öztürk, A discrete shrinking method as alternative to least-squares, *Communications de la Faculté des Sciences de l'Université d'Ankara* 33 (1984) 179–185.
- [18] S.D. Silvey, Multicollinearity and imprecise estimation, *J. Roy. Statist. Soc. B* 35 (1969) 67–75.
- [19] R.K. Smidt, L.L. McDonald, Ridge discriminant analysis, Research Paper No. 108, Stat. Lab S-1976-549, University of Wyoming, 1976.
- [20] A.N. Tikhonov, V.Y. Arsenin, *Solutions of Ill-posed Problems*, Wiley, New York, 1977.
- [21] G. Trenkler, An iteration estimator for the linear model, *Compstat, Physica Werlag, Würzburg*, 1978, 125–131.
- [22] H.D. Vinod, A. Ullah, *Recent Advances in Regression Methods*, Marcel Dekker, New York, 1981.
- [23] E. Walker, Detection of collinearity-influential observations, *Comm. Statist. Theory Methods* 18 (1989) 1675–1690.