mathematics

MDPI

Article

Solving High-Dimensional Problems in Statistical Modelling: A Comparative Study †

Stamatis Choudalakis ‡, Marilena Mitrouli ‡, Athanasios Polychronou ‡ and Paraskevi Roupa *,‡

Department of Mathematics, National and Kapodistrian University of Athens, Panepistimiopolis, 15784 Athens, Greece; stchoud@math.uoa.gr (S.C.); mmitroul@math.uoa.gr (M.M.); apolychronou@math.uoa.gr (A.P.)

* Correspondence: parask_roupa@math.uoa.gr

† This paper is dedicated to Mr. Constantin M. Petridi.

‡ These authors contributed equally to this work.

Abstract: In this work, we present numerical methods appropriate for parameter estimation in high-dimensional statistical modelling. The solution of these problems is not unique and a crucial question arises regarding the way that a solution can be found. A common choice is to keep the corresponding solution with the minimum norm. There are cases in which this solution is not adequate and regularisation techniques have to be considered. We classify specific cases for which regularisation is required or not. We present a thorough comparison among existing methods for both estimating the coefficients of the model which corresponds to design matrices with correlated covariates and for variable selection for supersaturated designs. An extensive analysis for the properties of design matrices with correlated covariates is given. Numerical results for simulated and real data are presented.

Keywords: high-dimensional; minimum norm solution; regularisation; Tikhonov; ℓ_p-ℓ_q; variable selection

Citation: Choudalakis, S.; Mitrouli, M.; Polychronou, A.; Roupa, P. Solving High-Dimensional Problems in Statistical Modelling: A Comparative Study. *Mathematics* **2021**, *9*, 1806. https://doi.org/10.3390/math9151806

Academic Editor: Nickolai Kosmatov

Received: 3 June 2021
Accepted: 26 July 2021
Published: 30 July 2021

1. Introduction

Many fields of science, and especially health studies, require the solution of problems in which the number of characteristics is larger than the sample size. These problems are referred to as high-dimensional problems. In the present paper, we focus on solving high-dimensional problems in statistical modelling.

We consider the linear regression model

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $X = \begin{bmatrix} \mathbf{1} & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix}$ is the design matrix of order $n \times (d+1)$, which is supposed to be high-dimensional, i.e., $n < d$. The columns $\boldsymbol{x}_i \sim N(\mathbf{0}_n, \sigma_i^2 I_n)$, $i = 1, 2, \ldots, d$, are the correlated covariates of the model and all the elements of the first column of the design matrix are equal to 1 in correspondence with the mean effect. The response vector $\boldsymbol{y}$ has length n, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$ is the n-vector of independent and identically distributed (i.i.d.) random errors, where $\epsilon_i \sim N(0, \sigma^2)$ for all $i = 1, 2, \ldots, n$.

In the present study we focus on the following two points.

1. Estimation of the regression parameter $\boldsymbol{\beta} \in \mathbb{R}^{d+1}$.
   From numerical linear algebra point of view, the statistical model (1) can be considered as an underdetermined system. This kind of system has infinitely many solutions. The first way to determine the desired vector $\boldsymbol{\beta}$ is to keep the solution with the minimum norm. This solution is referred to as minimum norm solution (MNS), [1] (p. 264). Another way of solving these problems is based on regularisation techniques. Specifically, these methods allow us to solve a different problem which

has a unique solution and thus to estimate the desired vector $\boldsymbol{\beta}$. One of the most popular regularisation methods is Tikhonov regularization, [2]. Another regularization technique which is used is the ℓ_p-ℓ_q regularization, [3,4].
It is of major importance to decide whether problem (1) can be solved directly in the least squares sense or regularisation is required. Therefore, we describe a way of choosing the appropriate method for solving (1) for design matrices with correlated covariates. For these matrices we study extensively their properties. We prove that as the correlation of the covariates increases, the generalised condition number of the design matrix increases as well and thus the design matrix becomes ill-conditioned.

2. To ascertain the most important factors of the statistical model.
Variable selection is a major issue in solving high-dimensional problems. By means of variable selection we refer to the specification of the important variables (active factors) in the linear regression model, i.e., the variables which play a crucial role in the model. The rest of the variables (inactive factors) can be omitted.
We deal with the variable selection in supersaturated designs (SSDs) which are fractional factorial designs in which the run size is less than the number of all the main effects. In this class of designs, the columns of X, except the first column, have elements ± 1. The symbols 1 and -1 are usually utilised to denote the high and low level of each factor, respectively. The correlation of SSDs is usually small, i.e., $r \leq 0.5$. The analysis of SSDs is a main issue in Statistics. Many methods for analysing these designs have been proposed. In [5], a Dantzig selector was introduced. Recently, a sure independence screening method has been applied in a model selection method in SSDs [6], and a support vector machine recursive feature elimination method for feature selection [7]. In our study, as we want to retain sparsity in variable selection, we adopt the ℓ_p-ℓ_q regularisation and the SVD principal regression method, [8], in order to determine the most important factors of the statistical model.

In the regression model (1), there is no error setting in the design matrix X which defines the model. It is always considered an unperturbed matrix X with covariates from normal distribution with well determined rank. However, we assume i.i.d. random error $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$, $\epsilon_i \sim N(0, \sigma^2)$ for all $i = 1, 2, \ldots, n$, incorporated in the model as given from relation (1). Thus, we are having well-posed problems on the set of the data according to the work in [9].

The paper is organised as follows. In Section 2, we briefly present some methods for solving high-dimensional problems. We initially display the MNS and in the sequel we present two regularisation methods. Specifically, Tikhonov regularisation and a general regularisation technique, ℓ_p-ℓ_q regularisation method, are discussed. The described methods are used in estimating the regression parameter $\boldsymbol{\beta}$ of (1) for design matrices with correlated covariates and the results are given in Section 3. These methods can be applied to ill-posed problems as well. Variable selection for SSDs can be found in Section 4. We end up this work with several concluding remarks in Section 5.

2. Methods Overview

In this section, we present some methods for solving high-dimensional problems.

2.1. Minimum Norm Solution

The system (1), which is an underdetermined system, does not have a unique solution. In fact, this underdetermined system has infinitely many solutions, and we are seeking a solution such that its norm is minimised, i.e., the minimum norm solution (MNS) $\underset{\boldsymbol{\beta} \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2$, [1] (p. 264). A necessary and sufficient condition for the existence of MNS is given in the following theorem.

Theorem 1. *Let $X \in \mathbb{R}^{n\times(d+1)}$ be a high-dimensional matrix, i.e., $n < d$, with $rank(X) = n$, and $\boldsymbol{\beta}^*$ be a solution of the underdetermined system $X\boldsymbol{\beta} = \boldsymbol{y}$. Then, $\boldsymbol{\beta}^*$ is a MNS if and only if $\boldsymbol{\beta}^* \in Range(X^T)$.*

Proof. As $\boldsymbol{\beta}^*$ is a solution of the underdetermined system $X\boldsymbol{\beta} = \boldsymbol{y}$, we have

$$X\boldsymbol{\beta}^* = \boldsymbol{y} \;\Leftrightarrow\; (\boldsymbol{\beta}^*)^T X^T = \boldsymbol{y}^T. \tag{2}$$

Let us consider the QR factorisation of X^T, i.e.,

$$X^T = QR = Q\begin{bmatrix} R_1 \\ 0_{d+1-n,n} \end{bmatrix},$$

where $Q \in \mathbb{R}^{(d+1)\times(d+1)}$ is orthogonal and $R_1 \in \mathbb{R}^{n\times n}$ is upper triangular. Therefore, (2) can be rewritten as

$$(\boldsymbol{\beta}^*)^T QR = \boldsymbol{y}^T \;\Leftrightarrow\; (Q^T\boldsymbol{\beta}^*)^T R = \boldsymbol{y}^T. \tag{3}$$

If we set

$$\boldsymbol{z} = Q^T\boldsymbol{\beta}^*, \tag{4}$$

then

$$(3) \;\Leftrightarrow\; \boldsymbol{z}^T R = \boldsymbol{y}^T \;\Leftrightarrow\; R^T\boldsymbol{z} = \boldsymbol{y}.$$

Moreover, we have

$$(4) \;\Leftrightarrow\; Q^{-1}\boldsymbol{\beta}^* = \boldsymbol{z} \;\Leftrightarrow\; \boldsymbol{\beta}^* = Q\boldsymbol{z} \;\Leftrightarrow\; \boldsymbol{\beta}^* \in Range(Q) \;\Leftrightarrow\; \boldsymbol{\beta}^* \in Range(X^T).$$

□

Taking into account the result of the above theorem, we obtain the formula for the MNS $\boldsymbol{\beta}^*$, which is given by

$$\boldsymbol{\beta}^* = X^T(XX^T)^{-1}\boldsymbol{y}. \tag{5}$$

Formula (5) cannot be used directly for calculating the vector $\boldsymbol{\beta}$, as it is not a stable computation. Therefore, we state the Algorithm 1 for a stable way of calculating the MNS through the singular value decomposition (SVD) of the design matrix X, [1] (p. 265). The operation count for this algorithm is dominated by the computation of the SVD, which requires a cost of $\mathcal{O}(nd^2)$ flops.

Algorithm 1: Computation of MNS via *SVD*.

Inputs: Design matrix $X \in \mathbb{R}^{n\times(d+1)}$, $n < d$, $rank(X) = n$
Response vector $\boldsymbol{y} \in \mathbb{R}^n$
Output: MNS solution $\boldsymbol{\beta}^*$

– Compute the *SVD* of X, i.e., $X = USV^T = \sum_{i=1}^{n} s_i\boldsymbol{u}_i\boldsymbol{v}_i^T$

– Compute the solution $\boldsymbol{\beta}^* = \sum_{i=1}^{n} \frac{\boldsymbol{u}_i^T\boldsymbol{y}}{s_i}\boldsymbol{v}_i$

2.2. The Discrete Picard Condition

It is crucial to identify when problem (1) can be directly solved with a satisfactory MNS solution or different ways of handling the solution must be employed. In [10], a criterion for deciding whether a least squares problem can have a satisfactory direct solution or not is proposed. This criterion employs the SVD of the design matrix X and the discrete Picard condition as defined in [11,12]. Let $X = USV^T = \sum_{i=1}^{n} s_i\boldsymbol{u}_i\boldsymbol{v}_i^T$ be the SVD of X, where s_i are the singular values of X with corresponding left singular vectors $\boldsymbol{u}_i$ and right singular

vectors $\boldsymbol{v}_i$, $i = 1, 2, \ldots, n$. The discrete Picard condition ensures that the solution can be approximated by a regularised solution [13].

Definition 1 (The discrete Picard condition)**.** *The discrete Picard condition (DPC) requires that the ratio* $\dfrac{|c_i|}{s_i}$ *decreases to zero as* $i \to n$*, i.e.,*

$$\frac{|c_i|}{s_i} \to 0, \quad as \;\; i \to n,$$

where $c_i = \boldsymbol{u}_i^T \boldsymbol{y}$*. The DPC implies that the constants* $|c_i|$ *tend to zero faster than the singular values tend to zero.*

Example 1. *Let us now consider design matrices of order* 50×101*, their columns have same variance* σ^2 *and same correlation structure* r*. In particular, we test two design matrices* X *with* $(r, \sigma^2) = (0.9, 0.25)$ *and* $(r, \sigma^2) = (0.999, 1)$*. In Figure 1, we display the ratios* $\dfrac{|c_i|}{s_i}$ *and* $\dfrac{|\hat{c}_i|}{s_i}$*, which correspond to the noise-free and the noisy problem,* $c_i = \boldsymbol{u}_i^T \boldsymbol{y}$*,* $\hat{c}_i = \boldsymbol{u}_i^T \hat{\boldsymbol{y}}$*,* $i = 1, 2, \ldots, n$*,* $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\epsilon}$*. If the graphs are close enough the MNS is satisfactory; otherwise, regularisation techniques are necessary for deriving a good approximation of the desired vector* $\boldsymbol{\beta}$*. As we can see in Figure 1, the values of the depicted ratios are very close in the design matrix with* $r = 0.9$ *case whereas in the highly correlated matrix with* $r = 0.999$ *case the ratios differ. This implies that a regularisation method is necessary for the second case.*

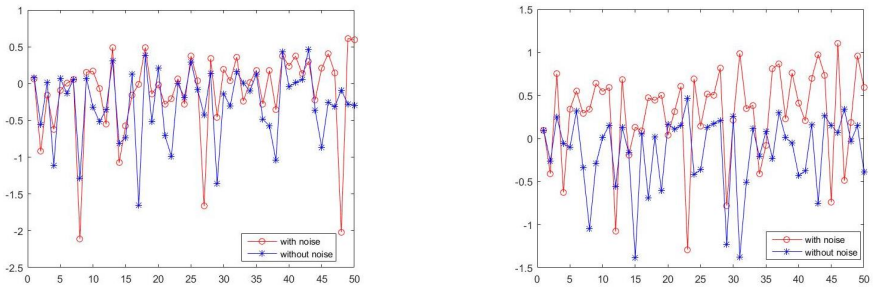

Figure 1. The ratios $|c_i|/s_i$ and $|\hat{c}_i|/s_i$ for the design matrices of order 50×101 for $(r, \sigma^2) = (0.9, 0.25)$ (**left**) and $(r, \sigma^2) = (0.999, 1)$ (**right**).

2.3. Regularisation Techniques

There are cases where the MNS $\boldsymbol{\beta}^*$ cannot achieve a good approximation of the desired unknown solution $\boldsymbol{\beta}$. As in the linear regression model as described in (1) the design matrix X is always unperturbed, and thus its rank can be a priori known, we can adopt regularisation techniques. In the present section, we present two regularisation methods. In particular, we present the popular Tikhonov regularisation [2] and the ℓ_p-ℓ_q regularisation which has recently received considerable attention [3,4]. Both of these techniques replace the initial problem with another one which is close to the original.

2.3.1. Tikhonov Regularisation

A regularisation method that is widely used is Tikhonov regularisation. The standard form of Tikhonov regularization, which corresponds in linear regression model (1), is given by

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{d+1}} \{ \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 + \lambda^2 \|\boldsymbol{\beta}\|_2^2 \}, \tag{6}$$

where λ is the regularisation parameter. The solution of the penalised least-squares problem (6) is given by the formula

$$\boldsymbol{\beta}_\lambda = (X^T X + \lambda^2 I_{d+1})^{-1} X^T \boldsymbol{y} = X^T (X X^T + \lambda^2 I_n)^{-1} \boldsymbol{y},$$

as it holds the identity $(X^TX + \lambda^2 I_{d+1})^{-1}X^T = X^T(XX^T + \lambda^2 I_n)^{-1}$. Indeed, we have $X^T(XX^T + \lambda^2 I_n) = (X^TX + \lambda^2 I_{d+1})X^T \Rightarrow (X^TX + \lambda^2 I_{d+1})^{-1}X^T = X^T(XX^T + \lambda^2 I_n)^{-1}$.

As we can see, Tikhonov regularisation depends on the regularization parameter λ. An appropriate method for selecting λ leads to the derivation of a satisfactory approximation $\boldsymbol{\beta}_\lambda$ of the desired regression parameter $\boldsymbol{\beta}$. The error $\boldsymbol{\epsilon}$ in the input data for the statistical model that we study follows the standard norm distribution, i.e., $\boldsymbol{\epsilon} \sim N(\mathbf{0}_n, \sigma^2 I_n)$. Therefore, the norm of the error is known, and it is given by $\|\boldsymbol{\epsilon}\|_2 = \sqrt{n-1}\sigma$. In the case of known error norm, the appropriate method for the selection of the regularisation parameter is the discrepancy principle, which is reported in Algorithm 2 [14] (p. 283). Following also the analysis presented in [9], and due to the uniqueness of λ for most reasonable values of ϵ (see, for example, in [15]), we adopt this method for our study.

Algorithm 2: Discrepancy principle.

Inputs: Design matrix $X \in \mathbb{R}^{n\times(d+1)}$, $n < d$, $rank(X) = n$
Response vector $\boldsymbol{y} \in \mathbb{R}^n$
Error norm $\|\boldsymbol{\epsilon}\|_2 = \sqrt{n-1}\sigma$
Output: Regularisation parameter λ
– Compute the SVD of X, i.e., $X = USV^T$
– Set $\boldsymbol{c} = U^T\boldsymbol{y}$
– Choose $\lambda > 0$ such that $\lambda^4 \boldsymbol{c}^T(S + \lambda^2 I)^{-2}\boldsymbol{c} = \|\boldsymbol{e}\|^2$, over a given grid of λ.

2.3.2. ℓ_p-ℓ_q Regularisation

A more general regularisation technique is the so-called ℓ_p-ℓ_q regularisation [3]. The main idea of this approach is based on the replacement of the minimisation problem $\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2$ by an ℓ_p-ℓ_q minimisation problem of the form

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{d+1}} \{\frac{1}{p}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_p^p + \mu\frac{1}{q}\|\boldsymbol{\beta}\|_q^q\}, \tag{7}$$

where $\mu > 0$ is the regularisation parameter and $0 < p, q \leq 2$. The solution of the minimisation problem (7) is given by

$$\hat{\boldsymbol{\beta}}_\mu = \underset{\boldsymbol{\beta}\in\mathbb{R}^{d+1}}{\operatorname{argmin}}\{\frac{1}{p}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_p^p + \mu\frac{1}{q}\|\boldsymbol{\beta}\|_q^q\}. \tag{8}$$

Remark 1. *In case of $p = q = 2$, the regularised minimisation problem (7) reduces to Tikhonov regularisation.*

Concerning the selection of the regularisation parameter, we choose the optimal value of μ, i.e., the value that minimises the error norm $\|\hat{\boldsymbol{\beta}}_\mu - \boldsymbol{\beta}\|_2$ over a given grid of values for μ. Concerning the computational cost, the implementation of the ℓ_p-ℓ_q regularisation requires $\mathcal{O}(nd)$ flops.

3. Design Matrix with Correlated Covariates

In high-dimensional applications, the design matrix $X = \begin{bmatrix} \mathbf{1} & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix}$ has correlated covariates $\boldsymbol{x_i} \sim N(\mathbf{0}_n, \sigma_i^2 I_n)$, $i = 1, \ldots, d$, where σ_i^2 is the variance of $\boldsymbol{x_i}$ and the correlation structure is given from the relation

$$r_{ij} = cor(\boldsymbol{x_i}, \boldsymbol{x_j}) = \frac{\boldsymbol{x_i}^T\boldsymbol{x_j}}{\|\boldsymbol{x_i}\|\|\boldsymbol{x_j}\|},\ i, j = 1, \ldots, d,\ i \neq j,$$

with $-1 \leq r_{ij} \leq 1$.

Next, we present a thorough investigation of the properties that characterize these matrices.

3.1. Correlated Covariates with Same Variance and Correlation

We initially consider design matrices with correlated covariates which have same variance σ^2 and same correlation r. In the following theorem, we formulate and prove in detail the types for the singular values of the design matrix X. In [16], this case of design matrix is considered and there exists a brief description of the eigenvalues of the matrix X^TX.

Theorem 2. *Let* $X = \begin{bmatrix} \mathbf{1} & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix} \in \mathbb{R}^{n\times(d+1)}$ *be a high-dimensional design matrix of full rank whose columns* $\boldsymbol{x}_i \sim N(\mathbf{0}_n, \sigma^2 I_n)$, $i = 1, 2, \ldots, d$, *with correlation structure* r. *The singular values of the matrix* X *are*

$$s_1 = \sqrt{n},\ \ s_2 = \cdots = s_{n-1} = \sigma\sqrt{(n-1)(1-r)},\ \ s_n = \sigma\sqrt{(n-1)[(d-1)r+1]}.$$

Proof. The n singular values of X are the square roots of the n non-zero eigenvalues of X^TX. Therefore, we compute the matrix X^TX, i.e.,

$$X^TX = \begin{bmatrix} 1 & \ldots & 1 \\ x_{11} & \ldots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1d} & \ldots & x_{nd} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & \ldots & x_{1d} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nd} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{n} 1 & \sum_{j=1}^{n} x_{j1} & \cdots & \sum_{j=1}^{n} x_{jd} \\ \sum_{j=1}^{n} x_{j1} & & & \\ \vdots & & \hat{X}^T\hat{X} & \\ \sum_{j=1}^{n} x_{jd} & & & \end{bmatrix} = \begin{bmatrix} n & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & \hat{X}^T\hat{X} & \\ 0 & & & \end{bmatrix},$$

where $\hat{X} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix}$ and $\sum_{j=1}^{n} x_{ji} = 0$, $\forall\, i = 1, \ldots, d$, due to the construction of the design matrix X according to the normal distribution. Therefore, the matrix X^TX has one eigenvalue equal to n.

Moreover, we can express the variance σ^2 of each covariate $\boldsymbol{x}_i = \begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{ni} \end{bmatrix}^T$ in terms of vector norms as follows:

$$\sigma^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{\boldsymbol{x}}_i)^2 = \frac{1}{n-1}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i\|^2,$$

where $\bar{\boldsymbol{x}}_i$ denotes the mean value of each $\boldsymbol{x}_i$. As the mean value of each $\boldsymbol{x}_i$ is zero, we have

$$\sigma^2 = \frac{1}{n-1}\|\boldsymbol{x_i}\|^2 \Rightarrow \|\boldsymbol{x_i}\|^2 = (n-1)\sigma^2,\ \ \forall\, i = 1, \ldots, d. \tag{9}$$

The submatrix $\hat{X}^T\hat{X}$ of X^TX can be written as

$$\hat{X}^T\hat{X} = \begin{bmatrix} \|\boldsymbol{x}_1\|^2 & r\|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\| & \dots & r\|\boldsymbol{x}_1\|\|\boldsymbol{x}_d\| \\ r\|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\| & \|\boldsymbol{x}_2\|^2 & \dots & r\|\boldsymbol{x}_2\|\|\boldsymbol{x}_d\| \\ \vdots & \vdots & \ddots & \vdots \\ r\|\boldsymbol{x}_d\|\|\boldsymbol{x}_1\| & r\|\boldsymbol{x}_d\|\|\boldsymbol{x}_2\| & \dots & \|\boldsymbol{x}_d\|^2 \end{bmatrix}$$

$$\overset{(9)}{=} \begin{bmatrix} (n-1)\sigma^2 & r(n-1)\sigma^2 & \dots & r(n-1)\sigma^2 \\ r(n-1)\sigma^2 & (n-1)\sigma^2 & \dots & r(n-1)\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ r(n-1)\sigma^2 & r(n-1)\sigma^2 & \dots & (n-1)\sigma^2 \end{bmatrix}$$

$$= (n-1)\sigma^2 \begin{bmatrix} 1 & r & \dots & r \\ r & 1 & \dots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \dots & 1 \end{bmatrix} = (n-1)\sigma^2[(1-r)I + rJ],$$

where J is the $d \times d$ matrix with all elements equal to 1. The non-zero eigenvalues of $\hat{X}^T\hat{X}$ are $\lambda_1 = (n-1)\sigma^2(1-r)$ with algebraic multiplicity $n-2$ and $\lambda_2 = (n-1)\sigma^2[(d-1)r+1]$ with algebraic multiplicity 1. Therefore, the singular values of X are $s_1 = \sqrt{n}$, $s_2 = \cdots = s_{n-1} = \sigma\sqrt{(n-1)(1-r)}$, $s_n = \sigma\sqrt{(n-1)[(d-1)r+1]}$. □

Let us denote by $\kappa(X)$ the generalised condition number of X, i.e., $\kappa(X) = \|X\|_2 \cdot \|X^\dagger\|_2$, where $X^\dagger = X^T(XX^T)^{-1}$ is the pseudoinverse of X, [1] (p. 246). It is known that the generalised condition number can be expressed in terms of the maximum s_{max} and the minimum s_{min} singular value of X as $\kappa(X) = \frac{s_{max}}{s_{min}}$, [1] (p. 216).

In Theorem 3, we express the generalised condition number of X in terms of the correlation structure r.

Theorem 3. *Let* $X = \begin{bmatrix} \mathbf{1} & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix} \in \mathbb{R}^{n\times(d+1)}$ *be a high-dimensional design matrix of full rank whose columns* $\boldsymbol{x}_i \sim N(\mathbf{0}_n, \sigma^2 I_n)$, $i = 1, 2, \dots, d$, *with correlation structure* r. *The generalised condition number of* X *is given by*

1. $\kappa(X) = \sqrt{\frac{n}{(n-1)\sigma^2(1-r)}}$, *if* $r \le \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right)$,
2. $\kappa(X) = \sqrt{\frac{(d-1)r+1}{1-r}}$, *if* $\left(r > \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right)$ *and* $\sigma^2 < \frac{n}{n-1}\right)$ *or* $\left(r > 1 - \frac{n}{(n-1)\sigma^2}$ *and* $\sigma^2 > \frac{n}{n-1}\right)$,
3. $\kappa(X) = \sqrt{\frac{(n-1)\sigma^2((d-1)r+1)}{n}}$, *if* $r < 1 - \frac{n}{(n-1)\sigma^2}$.

Proof. It is obvious that $s_n = \sigma\sqrt{(n-1)[(d-1)r+1]} > s_i$, $i = 2, \dots, n-1$ holds. Therefore, we have to distinguish three cases. The first case is $s_1 \ge s_n$, the second case is $s_i < s_1 < s_n$ and the last one is $s_1 < s_i$.

First case: If $s_1 \ge s_n$, then $\kappa(X) = \frac{s_1}{s_i} = \sqrt{\frac{n}{(n-1)\sigma^2(1-r)}}$. The restriction $s_1 \ge s_n$ can be rewritten as follows:

$$\begin{aligned} & n \geq (n-1)\sigma^2((d-1)r+1) \\ \Leftrightarrow \quad & \frac{n}{(n-1)\sigma^2} \geq (d-1)r+1 \\ \Leftrightarrow \quad & \frac{n}{(n-1)\sigma^2} - 1 \geq (d-1)r \\ \Leftrightarrow \quad & r \leq \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right). \end{aligned}$$

Second case: If $s_i < s_1 < s_n$, then $\kappa(X) = \frac{s_n}{s_i} = \sqrt{\frac{(d-1)r+1}{1-r}}$. The restriction $s_i < s_1 < s_n$ can be reformulated as follows:

$$\begin{aligned} & (n-1)\sigma^2(1-r) < n < (n-1)\sigma^2(dr+1-r) \\ \Leftrightarrow \quad & \begin{cases} 1-r < \frac{n}{(n-1)\sigma^2} \\ (d-1)r > \frac{n}{(n-1)\sigma^2} - 1 \end{cases} \quad \Leftrightarrow \begin{cases} r > 1 - \frac{n}{(n-1)\sigma^2} \\ r > \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right) \end{cases} . \end{aligned}$$

Moreover, we make the check

$$\begin{aligned} & \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right) < 1 - \frac{n}{(n-1)\sigma^2} \\ \Leftrightarrow \quad & n - (n-1)\sigma^2 < (d-1)(n-1)\sigma^2 - n(d-1) \\ \Leftrightarrow \quad & (n-1)\sigma^2(1+d-1) > n + nd - n \\ \Leftrightarrow \quad & \sigma^2 > \frac{n}{n-1}. \end{aligned}$$

Therefore, we conclude that the generalised condition number $\kappa(X)$ is equal to $\sqrt{\frac{(d-1)r+1}{1-r}}$ if the following relation holds.

$$\begin{aligned} r > \frac{1}{d-1}\left(\frac{n}{(n-1)\sigma^2} - 1\right) \quad & \text{and} \quad \sigma^2 < \frac{n}{n-1} \quad \text{or} \\ r > 1 - \frac{n}{(n-1)\sigma^2} \quad & \text{and} \quad \sigma^2 > \frac{n}{n-1} \end{aligned}$$

Third case: If $s_1 \leq s_i$, then $\kappa(X) = \frac{s_n}{s_1} = \sqrt{\frac{(n-1)\sigma^2((d-1)r+1)}{n}}$. This restriction is equivalently written as

$$n < (n-1)\sigma^2(1-r) \Leftrightarrow \frac{n}{(n-1)\sigma^2} < 1 - r \Leftrightarrow r < 1 - \frac{n}{(n-1)\sigma^2}.$$

□

Taking into consideration the derived formulae for the generalised condition number of the design matrix X, we see that if $r \approx 1$ the generalised condition number $\kappa(X)$ becomes large. A detailed example is presented next.

Example 2. *In this example, we plot the generalised condition number of X as a function of the correlation r. We consider* $n = 50$, $d = 100$ *and* $\sigma^2 = 2$. *In Figure 2, we display* $\kappa(X)$ *for correlation* $r \longrightarrow 1$. *As we see in Figure 2, as correlation r tends to* 1, *the generalised condition number* $\kappa(X)$ *increases rapidly.*

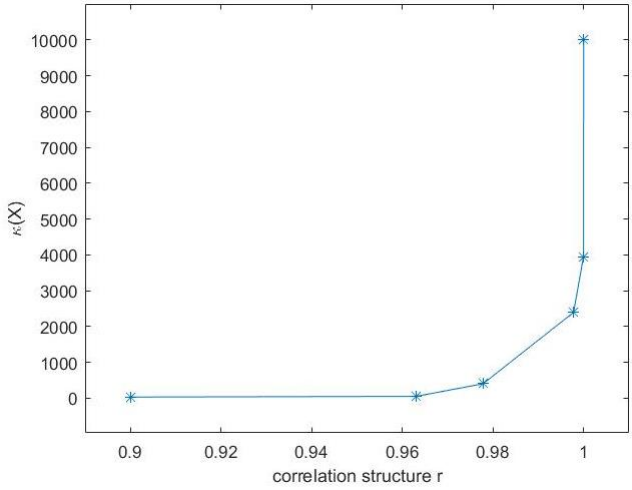

Figure 2. The generalised condition number of X as a function of the correlation r.

3.2. Highly Correlated Covariates with Different Variance and Correlation

Next, we consider a general and more usual case in which the covariates $\boldsymbol{x_i} \sim N(\mathbf{0}_n, \sigma_i^2 I_n)$ of the design matrix X have different variance σ_i^2 and correlation r_{ij}, $i, j = 1, \ldots, d$. Based on the results presented in [17] for the eigenvalues of the matrix $X^T X$, we record analytic formulae for the singular values of X in the following theorem.

Theorem 4. *Let* $X = \begin{bmatrix} \mathbf{1} & \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_d \end{bmatrix} \in \mathbb{R}^{n \times (d+1)}$ *be a high-dimensional design matrix of full rank whose columns* $\boldsymbol{x}_i \sim N(\mathbf{0}_n, \sigma_i^2 I_n)$, $i = 1, 2, \ldots, d$, *with highly correlation structure* r_{ij}. *The singular values of the matrix* X *are*

$$s_1 = \sqrt{n}, \;\; s_2 = \sqrt{(n-1)\sum_{j=1}^{d} \sigma_i^2 + \mathcal{O}(\delta)}, \;\; s_3 = \cdots = s_n = \sqrt{\mathcal{O}(\delta)},$$

assuming that $1 - r_{ij} = \mathcal{O}(\delta)$ *as* $\delta \to 0$.

As we record in Section 3.1, the generalised condition number is equal to the ratio $\frac{s_{max}}{s_{min}}$ and in the present case $s_{min} = \mathcal{O}(\delta)$ considering that $1 - r_{ij} = \mathcal{O}(\delta)$, i.e., highly correlated covariates. Therefore, the value of $\kappa(X)$ is large and this affects the solution of the corresponding problem.

Remark 2. *As the correlation* r *increases the generalised condition number* $\kappa(X)$ *increases as well. From Theorems 2 and 4 we deduce that the case of highly correlated covariates leads to possible instability and thus regularisation is recommended. This result is confirmed from Table 1 which is presented in Section 3.3.*

3.3. Numerical Implementation

The implementation of the simulation study presented in this section and in Section 4 has been done by using the Julia Programming Language.

Given the high-dimensional design matrix X of order $n \times (d+1)$, the response vector $\boldsymbol{y}$ of order n and the n-vector $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$ of i.i.d. random errors, $\epsilon_j \sim N(0,1)$, $j = 1, 2, \ldots, n$, we estimate the vector $\boldsymbol{\beta}$ by using the methods which are described in Section 2. We consider design matrices X with correlated covariates and we distinguish the two aforementioned cases. The results for the first case, i.e., the covariates of the design matrices having same correlation r and same variance σ^2, are recorded in Tables 1 and 2. The results for the second case are displayed in Table 3.

The implemented simulation scheme is the following. For each design matrix X, a random vector $\boldsymbol{\beta}$ is generated and $\boldsymbol{y} = X\boldsymbol{\beta}$ denotes the noise free response vector. Then,

100 iterations are performed, in each one the response vector is perturbed by noise $\boldsymbol{\epsilon}^i$ resulting in a noisy response vector $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\epsilon}^i$, $i = 1, 2, \ldots, 100$. Eventually, the regression parameter $\hat{\boldsymbol{\beta}}$ is computed by using both the MNS given by Algorithm 1 and the regularisation techniques. The quality of the generated approximation solution $\hat{\boldsymbol{\beta}}$ is assessed by the mean square error (MSE) between $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$ which is given by the formula

$$MSE(\hat{\boldsymbol{\beta}}) = E[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2].$$

In Algorithm 3, we summarise the simulation scheme.

Algorithm 3: Simulation scheme.

Input: Design matrix $X \in \mathbb{R}^{n\times(d+1)}$
Result: $MSE(\hat{\boldsymbol{\beta}})$
$\boldsymbol{\beta} = randn(n)$;
$\boldsymbol{y} = X\boldsymbol{\beta}$;
for $i \leftarrow 1$ **to** 100 **do**
  $\hat{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{\epsilon}^i$;
  $\boldsymbol{\beta}^* = MNS(\hat{\boldsymbol{y}})$;
  $\boldsymbol{\beta}_\lambda = Tikhonov(\hat{\boldsymbol{y}})$;
  $\hat{\boldsymbol{\beta}}_\mu = \ell_p$-$\ell_q(\hat{\boldsymbol{y}})$
end

In Tables 1–3, we present the results of estimating the regression parameter $\boldsymbol{\beta}$ for different orders of the design matrices X. In the two first columns of the tables, the correlation r and the variance σ^2 of the covariates are recorded, respectively. In Table 3, we record the interval in which lies the correlation and the variance. In the third column, the adopted methods are written. Specifically, we record MNS, Tikhonov regularisation and ℓ_p-ℓ_q regularisation technique for different pairs of (p, q). The fourth column contains the used grid of values for the regularisation parameter λ or μ for Tikhonov or ℓ_p-ℓ_q regularisation, respectively. In the last column the $MSE(\hat{\boldsymbol{\beta}})$ of the derived approximation solutions $\hat{\boldsymbol{\beta}}$ are recorded.

Table 1. Results for $X_{5\times 21}$.

r	σ^2	Method	λ/μ	MSE ($\hat{\boldsymbol{\beta}}$)
0.5	0.25	MNS		1.3063×10^{-1}
		Tikhonov	[1, 10]	8.3874×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	$[10^{-7}, 10^{-2}]$	1.1949×10^{-1}
0.5	1.0	MNS		1.3093×10^{-1}
		Tikhonov	[1, 10]	8.0127×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	$[10^{-7}, 10^{-2}]$	1.2216×10^{-1}
0.9	0.25	MNS		5.5782×10^{-1}
		Tikhonov	[1, 10]	9.4101×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	1.2571×10^{-1}
0.9	1.0	MNS		6.2096×10^{-1}
		Tikhonov	[1, 10]	8.5836×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	6.0884×10^{-1}
0.999	0.25	MNS		4.4474
		Tikhonov	[1, 10]	1.774
		$\ell_{0.1}$-ℓ_2	$[10^{-7}, 10^{-2}]$	7.3793×10^{-1}
0.999	1.0	MNS		2.0129
		Tikhonov	[1, 10]	1.0456
		$\ell_{1.2}$-$\ell_{1.2}$	[0.1, 10]	6.8626×10^{-1}

Table 2. Results for $X_{50\times101}$.

r	σ^2	Method	λ/μ	MSE ($\hat{\beta}$)
0.9	0.25	MNS		4.5894×10^{-1}
		Tikhonov	[1, 10]	7.0511×10^{-1}
		ℓ_2-$\ell_{0.1}$	$[10^{-7}, 10^{-2}]$	4.5093×10^{-1}
0.9	1.0	MNS		4.8802×10^{-1}
		Tikhonov	[1, 10]	6.0754×10^{-1}
		ℓ_2-$\ell_{0.1}$	$[10^{-7}, 10^{-2}]$	4.8614×10^{-1}
0.999	0.25	MNS		3.5306
		Tikhonov	[1, 10]	1.0022
		ℓ_2-$\ell_{0.1}$	[0.1, 10]	8.3247×10^{-1}
0.999	1.0	MNS		1.1970
		Tikhonov	[1, 10]	9.6625×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	7.4754×10^{-1}

Table 3. Results for $X_{25\times51}$.

r	σ^2	Method	λ/μ	MSE ($\hat{\beta}$)
[0.27, 0.91]	[0.19, 1.17]	MNS		2.1252×10^{-1}
		Tikhonov	[1, 10]	6.2245×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	9.3163×10^{-2}
[−0.32, 0.85]	[0.13, 2.32]	MNS		1.6819×10^{-1}
		Tikhonov	[1, 10]	5.9699×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	1.2632×10^{-1}
[0.06, 0.91]	[0.42, 1.93]	MNS		1.1623×10^{-1}
		Tikhonov	[1, 10]	5.8305×10^{-1}
		$\ell_{1.8}$-$\ell_{1.8}$	[0.1, 10]	1.0371×10^{-1}

As we can see in these tables, in the case of highly correlated design matrices, the regularisation is necessary for deriving a good approximation of the desired vector $\boldsymbol{\beta}$. On the other hand, if the correlation of the design matrix is not high, MNS can achieve a fair estimation and a regularisation method does not improve the results, as it is verified by the $MSE(\hat{\boldsymbol{\beta}})$. Therefore, according to the presented results, for matrices with moderate correlated covariates, regularisation is redundant, as MNS yields adequate results. However, as the correlation between the covariates rises, the regularisation is essential.

Note that in case of design matrices with same variance and correlation $r = 0.999$ (Tables 1 and 2) the regularisation techniques, Tikhonov and ℓ_p-ℓ_q, can achieve comparable results. The choice of the pair of parameters (p, q) and the values of the required regularisation parameter play an important role for the efficient implementation of both methods.

4. Variable Selection in SSDs

In this section, we are interested in selecting the active factors of SSDs by using the methods which are described in Section 2. In our comparison, we also include SVD principal regression method which is used in SSDs, and it was proposed in [8]. We briefly refer to this method as SVD regression. The main computational cost of this approach is the evaluation of the SVD.

We measure the effectiveness of these methods through the Type I and Type II error rates. In particular, Type I error measures the cost of declaring an inactive factor to be active and Type II measures the cost of declaring an active effect to be inactive. In our numerical experiments, we consider 500 different realisations of the error $\boldsymbol{\epsilon}$ and in the presented tables we record the mean value of Type I, II error rates.

It is worth mentioning that both the MNS and Tikhonov regularisation give that all the factors are active, i.e., Type I $= 1$, Type II $= 0$, for all the tested SSDs. Therefore, these

methods are not suitable for variable selection and we do not include them in the following presented tables.

Example 3 (An illustrative example)**.** *In this example, we shall exhibit in detail the performance of each method for a particular problem. For this purpose, we adopt the illustrative example presented in [8], with design matrix*

$$X = \begin{bmatrix} + & - & - & - & - & - & - & - & - & - \\ + & + & + & + & + & + & + & - & - & - \\ + & + & + & + & - & - & - & + & + & + \\ - & + & - & - & + & + & - & + & + & - \\ - & - & + & - & + & - & + & + & - & + \\ - & - & - & + & - & + & + & - & + & + \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_{10} \end{bmatrix}.$$

Then a first column $\boldsymbol{x}_0$ *with all entries equal to* 1 *is added to the matrix, which corresponds to the average mean. The simulated data are generated by the model*

$$\boldsymbol{y} = 5\boldsymbol{x}_0 + 4\boldsymbol{x}_2 + 3\boldsymbol{x}_5 + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}_6, I_6)$*. A response vector* $\boldsymbol{y}$ *obtained by using this model is*

$$\boldsymbol{y} = \begin{bmatrix} -1.54 & 12.02 & 6.82 & 12.44 & 4.62 & -1.21 \end{bmatrix}^T.$$

The exact regression parameter $\boldsymbol{\beta}$ *and the predicted coefficients by each method are demonstrated below.*

$$\begin{aligned}
\boldsymbol{\beta} &= \begin{bmatrix} 5 & 0 & 4 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T, \\
\hat{\boldsymbol{\beta}}_{MNS} &= \begin{bmatrix} 5.525 & 0.1208 & 2.4508 & 1.1475 & 0.1758 & 2.0842 & 1.1125 & -0.1908 & 1.2175 & 0.2458 & -1.0575 \end{bmatrix}^T, \\
\hat{\boldsymbol{\beta}}_{Tik} &= \begin{bmatrix} 4.7237 & 0.1114 & 2.2592 & 1.0578 & 0.1621 & 1.9212 & 1.0255 & -0.1759 & 1.1223 & 0.2266 & -0.9748 \end{bmatrix}^T, \\
\hat{\boldsymbol{\beta}}_{\ell_2\ell_{0.1}} &= \begin{bmatrix} 5.5214 & 0.0 & 3.9482 & 0.0 & 0.0 & 2.8458 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}^T, \\
\hat{\boldsymbol{\beta}}_{SVD} &= \begin{bmatrix} 5.5245 & 0.0 & 3.901 & 0.0 & 0.0 & 2.801 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}^T.
\end{aligned}$$

As we can see from the generated approximation solutions $\hat{\boldsymbol{\beta}}$, the MNS and Tikhonov regularised solution cannot specify the active factors of the model and completely spoil the sparsity. On the other hand, the ℓ_p-ℓ_q regularisation method and the SVD regression can determine appropriately the active factors of the model.

Example 4 (Williams' data)**.** *We consider the well-known Williams' dataset (rubber age data) which is reported in Table 4. It is a classical dataset of SSDs and it is tested in several works, such as in [8]. As it is written in [8], as the columns* 13 *and* 16 *in the original design matrix are identical, the column* 13 *is removed for executing our numerical experiments. For this dataset we consider two cases, the real case and* 3 *synthetic cases.*

We initially deal with the real case where the design matrix X and the response vector $\boldsymbol{y}$ are given, without the initial knowledge of the desired vector $\boldsymbol{\beta}$. In literature, it is reported that the active factor is $\boldsymbol{x}_{15}$. In this case, according to our numerical experiments, the SVD regression and the ℓ_p-ℓ_q regularisation method for $p = 0.8$, $q = 0.1$ indicate that the factor $\boldsymbol{x}_{15}$ is important. In particular, the proposed models, i.e., the coefficients β_i are given in Table 5.

The second case corresponds to 3 synthetic cases, see in [8] and references therein, which are given below. For these simulated cases, we record the results in Table 6. In particular, we compute Type I and II error rates for the described methods. We apply the ℓ_p-ℓ_q regularisation method for $\mu = 5$ and the SVD regression for the significance level $a = 0.05$. As we notice in this table, both the ℓ_p-ℓ_q regularisation and the SVD regression can select sufficiently the important factors, as we see from the corresponding Type I, II error rates. The first model has the particularity that it includes the interaction of the factors

x_5, x_9 which does not usually appear in SSDs analysis. The first model is a challenging case for all the methods.

Model 1: $\boldsymbol{y} \sim N(15\boldsymbol{x}_1 + 8\boldsymbol{x}_5 - 6\boldsymbol{x}_9 + 3\boldsymbol{x}_5\boldsymbol{x}_9, I_{14})$
Model 2: $\boldsymbol{y} \sim N(8\boldsymbol{x}_1 + 5\boldsymbol{x}_{12}, I_{14})$
Model 3: $\boldsymbol{y} \sim N(10\boldsymbol{x}_1 + 9\boldsymbol{x}_2 + 2\boldsymbol{x}_3, I_{14})$

Table 4. The Williams' data—rubber age data.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	y
+	+	+	-	-	-	+	+	+	+	+	-	-	-	+	+	-	-	+	-	-	-	+	133
+	-	-	-	-	-	+	+	+	-	-	-	+	+	+	-	+	-	-	+	+	-	-	62
+	+	-	+	+	-	-	-	-	+	-	+	+	+	+	+	-	-	-	-	+	+	-	45
+	+	-	+	-	+	-	-	-	+	+	-	-	+	+	-	+	+	+	-	-	-	-	52
-	-	+	+	+	+	-	+	+	-	-	-	-	+	+	+	-	-	+	-	+	+	+	56
-	-	+	+	+	+	+	-	+	+	+	-	+	+	-	+	+	+	+	+	+	-	-	47
-	-	-	-	+	-	-	+	-	+	-	+	+	-	+	+	+	+	+	+	-	-	+	88
-	+	+	-	-	+	-	+	-	+	-	-	-	-	-	-	-	+	-	+	+	+	-	193
-	-	-	-	-	+	+	-	-	-	+	+	-	+	-	+	+	-	-	-	-	+	+	32
+	+	+	+	-	+	+	+	-	-	-	+	+	+	-	+	-	+	-	+	-	-	+	53
-	+	-	+	+	-	-	+	+	-	+	-	+	-	-	-	+	+	-	-	-	+	+	276
+	-	-	-	+	+	+	-	+	+	+	+	-	-	+	-	-	+	-	+	+	+	+	145
+	+	+	+	+	-	+	-	+	-	-	+	-	-	-	+	-	+	+	-	+	-	-	130
-	-	+	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	+	-	+	-	-	127

Table 5. The selected model for William's data (real case).

Method	Intercept	x_{15}
$\ell_{0.8}$-$\ell_{0.1}$	6.11	−1.13
SVD Regression	102.7857	−36.0341

Table 6. Results for William's Data (synthetic cases).

Model	Method	Type I	Type II
Model 1	$\ell_{1.8}$-$\ell_{0.8}$	0.23	0.56
	SVD Regression	0.15	0.74
Model 2	$\ell_{2.0}$-$\ell_{0.1}$	0.00	0.00
	SVD Regression	0.05	0.00
Model 3	$\ell_{2.0}$-$\ell_{0.1}$	0.00	0.27
	SVD Regression	0.07	0.33

Example 5 (A 3-circulant SSD)**.** *In this example, we consider one more SSD, which is also used in [18], and it is recorded in Table 7. We test the behaviour of the methods for variable selection by considering three models which can be found in [19] and are given below.*

Model 1: $\boldsymbol{y} \sim N(10\boldsymbol{x}_1, I_8)$
Model 2: $\boldsymbol{y} \sim N(-15\boldsymbol{x}_1 + 8\boldsymbol{x}_5 - 2\boldsymbol{x}_9, I_8)$
Model 3: $\boldsymbol{y} \sim N(-15\boldsymbol{x}_1 + 12\boldsymbol{x}_5 - 8\boldsymbol{x}_9 + 6\boldsymbol{x}_{13} - 2\boldsymbol{x}_{17}, I_8)$

The results are presented in Table 8. For the three used models, we apply the ℓ_p-ℓ_q regularisation method for $\mu = 5, 5.5, 0.5$ respectively and the SVD regression for $a = 0.25$. According to the presented numerical results, we see that both the ℓ_p-ℓ_q regularisation and the SVD regression can achieve satisfactory Type I and II error rates for the Model 1. On the other hand, for the Model 2 the SVD regression fails to specify the active factors whereas the ℓ_p-ℓ_q regularisation method achieves better Type II error. However, neither of the methods produce fair results for the Model 3. The coefficients of this model are not sufficiently close and this fact affects the behaviour of the methods.

Table 7. A 3-circulant SSD.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
-	-	-	-	-	-	-	+	+	-	-	-	+	-	+	+	+	-	+	+	+
+	+	+	-	-	-	-	-	-	-	+	+	-	-	-	+	-	+	+	+	-
+	+	-	+	+	+	-	-	-	-	-	-	-	+	+	-	-	-	+	-	+
+	-	+	+	+	-	+	+	+	-	-	-	-	-	-	-	+	+	-	-	-
-	-	-	+	-	+	+	+	-	+	+	+	-	-	-	-	-	-	-	+	+
-	+	+	-	-	-	+	-	+	+	+	-	+	+	+	-	-	-	-	-	-
-	-	-	-	+	+	-	-	-	+	-	+	+	+	-	+	+	+	-	-	-
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 8. Results for 3-circulant SSD.

Model	Method	Type I	Type II
Model 1	$\ell_{2.0}$-$\ell_{0.1}$	0.00	0.00
	SVD Regression	0.08	0.00
Model 2	$\ell_{0.6}$-$\ell_{1.3}$	0.39	0.13
	SVD Regression	0.08	0.67
Model 3	$\ell_{1.8}$-$\ell_{0.8}$	0.39	0.45
	SVD Regression	0.17	0.80

Example 6. *In this example we consider a real data set presented in [20] that deals with moss bags of Rhynchostegium riparioides which were exposed to different water concentrations of* 11 *trace elements under laboratory conditions. The design matrix X can be found in Table 1 in [20]. We consider the main effects, the second- and third-order interactions of influent factors. Therefore, we have a* 67×232 *SSD and we can select the important factors applying the* ℓ_p-ℓ_q *regularisation for* $\mu = 0.75$ *and the SVD regression for significance level* $a = 0.05$.

From Table 9, we see that both ℓ_2-$\ell_{0.1}$ and SVD regression methods identify the main effect Zn as active factor. The second order interactions Cd/Mn, As/Pb and Mn/Ni are also identified as active. These results are in agreement with [20].

Table 9. Important elements and interactions.

Method	Main Effects	Second-Order Interactions	Third-Order Interactions
ℓ_2-$\ell_{0.1}$	Fe, Zn	Al/Hg, As/Pb, Cd/Mn, Mn/Ni	Al/As/Mn, Al/Cr/Zn, As/Cd/Fe, As/Cd/Mn, Cr/Mn/Zn, Cu/Hg/Mn Fe/Hg/Ni, Fe/ Ni/Pb
SVD Regression	Zn	As/Pb, Cd/Mn, Fe/Mn, Fe/Zn, Mn/Ni, Pb/Zn	

5. Conclusions

In the present work, we analysed the properties of design matrices with correlated covariates. Specifically, we derived and proved formulae for the singular values of these matrices and we studied the connection of the generalised condition number with the correlation structure. Moreover, we described some available methods for solving high-dimensional problems. We checked the behaviour of the MNS and the necessity of applying regularisation techniques in estimating the regression parameter $\boldsymbol{\beta}$ in the linear regression model. We concluded that in solving high-dimensional statistical problems the following remarks must be taken into consideration.

1. Regularisation should be applied only if the given data set satisfies the discrete Picard condition. In this case, the choice of the regularisation parameter can be uniquely chosen by applying the discrepancy principle method.
2. The regression parameter $\boldsymbol{\beta} \in \mathbb{R}^{d+1}$ can be satisfactory estimated by the MNS if the design matrix is not highly correlated but in case of highly correlated data matrices we have to adopt regularisation techniques. The quality of the derived estimation $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is assessed by the computation of $MSE(\hat{\boldsymbol{\beta}})$.
3. In variable selection, where sparse solutions are needed, SVD regression or ℓ_p-ℓ_q regularisation can be used. When only few factors of the experiment are needed to be specified (maybe only the most important), SVD regression may be preferable since it avoids regularisation and the troublesome procedure of defining the regularisation parameter. The quality of the variable selection which is proposed by the estimation methods is assessed by the evaluation of Type I and II error rates.

In conclusion, the proposed scheme for the selection of the appropriate method for the solution of high-dimensional statistical problems is summarised in the following logical diagram, see Figure 3.

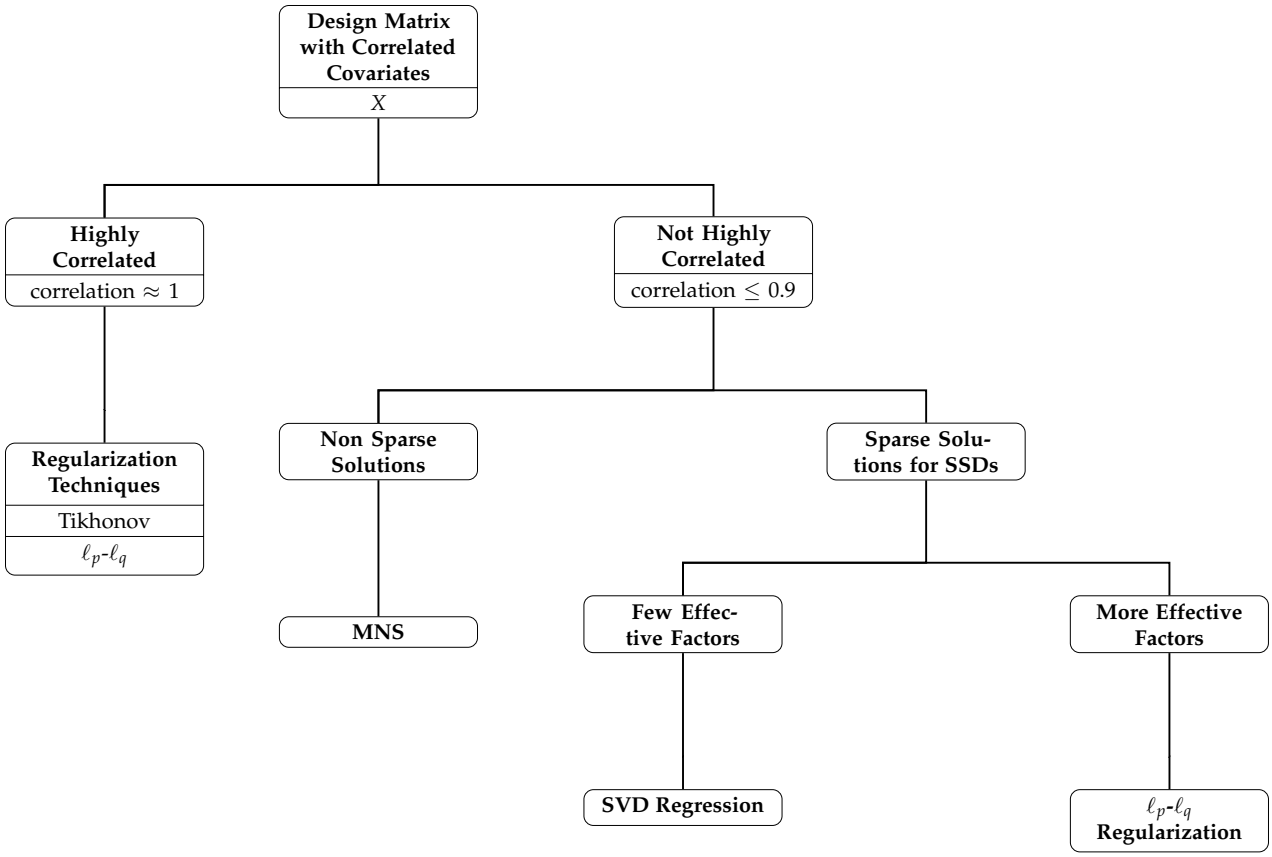

Figure 3. Logical diagram for choosing the appropriate method for the solution of high-dimensional statistical problems.

Author Contributions: Conceptualization, M.M.; methodology, M.M. and P.R.; software, S.C., A.P., P.R.; validation, M.M. and P.R.; formal analysis, S.C., A.P., P.R.; investigation, M.M.; data curation, S.C., A.P., P.R.; writing—original draft preparation, S.C., M.M., A.P., P.R.; writing—review and editing, S.C., M.M., A.P., P.R.; supervision, M.M.; project administration, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors are grateful to the reviewers of this paper whose valuable remarks improved this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Datta, B.N. *Numerical Linear Algebra and Applications*, 2nd ed.; SIAM: Philadelphia, PA, USA, 2010.
2. Tikhonov, A.N. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR* **1963**, *151*, 501–504.
3. Huang, G.; Lanza, A.; Morigi, S.; Reichel, L.; Sgallari, F. Majorization-minimization generalized Krylov subspace methods for ℓ_p-ℓ_q optimization applied to image restoration. *BIT Numer. Math.* **2017**, *57*, 351–378. [CrossRef]
4. Buccini, A.; Reichel, L. An ℓ_2-ℓ_q regularization method for large discrete ill-posed problems. *J. Sci. Comput.* **2019**, *78*, 1526–1549. [CrossRef]
5. Candes, E.; Tao, T. The Dantzig Selector: Statistical Estimation When p is much larger than n. *Ann. Stat.* **2007**, *35*, 2313–2351.
6. Drosou, K.; Koukouvinos, C. Sure independence screening for analyzing supersaturated designs. *Commun. Stat. Simul. Comput.* **2019**, *48*, 1979–1995. [CrossRef]
7. Drosou, K.; Koukouvinos, C. A new variable selection method based on SVM for analyzing supersaturated designs. *J. Qual. Technol.* **2019**, *51*, 21–36. [CrossRef]
8. Georgiou, S.D. Modelling by supersaturated designs. *Comput. Stat. Data Anal.* **2008**, *53*, 428–435. [CrossRef]
9. Yagola, A.G.; Leonov, A.S.; Titarenko, V.N. Data errors and an error estimation for ill-posed problems. *Inverse Probl. Eng.* **2002**, *10*, 117–129. [CrossRef]
10. Winkler, J.R.; Mitrouli, M. Condition estimation for regression and feature selection. *J. Comput. Appl. Math.* **2020**, *373*, 112212. [CrossRef]
11. Hansen, P.C. The discrete picard condition for discrete ill-posed problems. *BIT* **1990**, *30*, 658–672. [CrossRef]
12. Hansen, P.C. *Rank-Deficient and Discrete Ill-Posed Problems*; SIAM: Philadelphia, PA, USA, 1998.
13. Hansen, P.C. The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology, Advances in Computational Bioengineering 4*; WIT Press: Southampton, UK, 2000; pp. 119–142.
14. Golub, G.H.; Meurant, G. *Matrices, Moments and Quadrature with Applications*; Princeton University Press: Princeton, NJ, USA, 2010.
15. Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Kluwer: Dordrecht, The Netherlands, 1996.
16. Koukouvinos, C.; Lappa, A.; Mitrouli, M.; Roupa, P.; Turek, O. Numerical methods for estimating the tuning parameter in penalized least squares problems. *Commun. Stat. Simul. Comput.* **2019**, 1–22. [CrossRef]
17. Koukouvinos, C.; Jbilou, K.; Mitrouli, M.; Turek, O. An eigenvalue approach for estimating the generalized cross validation function for correlated matrices. *Electron. J. Linear Algebra* **2019**, *35*, 482–496. [CrossRef]
18. Liu, Y.; Dean, A. k-Circulant supersaturated designs. *Technometrics* **2004**, *46*, 32–43. [CrossRef]
19. Li, R.; Lin, D.K.J. Analysis Methods for Supersaturated Design: Some Comparisons. *J. Data Sci.* **2003**, *1*, 249–260. [CrossRef]
20. Cesa, M.; Campisi, B.; Bizzotto, A.; Ferraro, C.; Fumagalli, F.; Nimis, P.L. A Factor Influence Study of Trace Element Bioaccumulation in Moss Bags. *Arch. Environ. Contam. Toxicol.* **2008**, *55*, 386–396. [CrossRef] [PubMed]