

Regularised Least Squares estimation of external parameters

November 28, 2019

Preliminaries

Recall the notation from the previous reports:

$k = 1, \dots, K$ - index of the dataset used for identification.

$t = 1, \dots, T$ - discrete time index.

N_s - number of significant terms in the model.

$i = 1, \dots, N_s$ - term index.

$\Xi = \mathbb{R}^2$ - space of external parameters.

$\Theta = \mathbb{R}^{N_s}$ - space of internal parameters.

θ_i^k - an internal regression parameter corresponding to the i -th term in the k -th dataset.

$\theta^k = \{\theta_i^k\}_{i=1}^{N_s}$ - vector of internal parameters.

The analysis deals with the time-series data from auxetic foam vibration tests introduced in the previous report. A non-linear FIR filter is considered with the input lag of 3. The filter structure is approximated using polynomial regressors of order 2. Number the significant terms and corresponding scaling parameters θ are estimated via EFOR-CMSS algorithm based on 8 datasets out of 10 available. The remaining to datasets (C3 and C8) are used for model validation.

This report deals only with identification of the relationship between the estimated internal parameters θ^k and the pre-specified design settings $\xi^k = (L_{cut}, D_{rlx})$. The first section examines the effect on the polynomial structure and sampling from the small subspace of external parameters on the generalisation of the identified model. The remaining sections briefly describe approaches to improving the estimation procedure. The results section only demonstrates numerical results and does not draw any conclusions as this report is intermediate.

1 Collinearity in the selected model

Recall the assumed polynomial model that links internal parameters to the external ones:

$$\theta_i^k(L_{cut}, D_{rlx}) = \beta_0 + \beta_1 L_{cut} + \beta_2 D_{rlx} + \beta_3 L_{cut}^2 + \beta_4 D_{rlx}^2 + \beta_5 L_{cut} D_{rlx}, \quad i = 1, \dots, N_s, \quad (1)$$

where polynomial coefficients $\beta^i = [\beta_0 \dots \beta_5]$ are unknown. In the original paper [1] they are estimated via batch LS estimator as follows. First, the matrix of "independent" variables is formed from the external parameter values:

$$\mathbf{A} = \begin{bmatrix} 1 & L & D & L \times L & D \times D & L \times D \end{bmatrix}, \quad (2)$$

where $L = \{L_{cut}^k\}_{k=1}^K$, $D = \{D_{rlx}^k\}_{k=1}^K$, the number of columns $M + 1$, where M is the order of polynomial (1), and where \times denotes the by-element product such that

$$L \times L = \left[L_{cut}^{(j)} L_{cut}^{(j)} \right]_{j=1}^N.$$

Then, for each internal parameter a vector of estimated values is formed across K training datasets:

$$\bar{\theta}^i = \{\theta_i^1, \theta_i^1, \dots, \theta_i^K\}.$$

Note that while the internal parameters are considered to be fixed in the estimation, their estimates are obtained from the random data and therefore are random variables

$$\bar{\theta}^i \sim \mathcal{N}(\mu_i, \sigma_i^2).$$

Remark 1.1: *For the purposes of this analysis the factors contributing to this variance are not considered and the vector of LSEs is treated as a response signal in the LS formulation.*

Estimation of the polynomial coefficients for the i -th internal parameter β^i is a standard LS problem:

$$\hat{\beta}^i = \quad (3)$$

where

It can be seen from Table 1 in previous report that the experimental data is collected from the narrow subspace of design parameters. Combined with the polynomial structure that means that the columns of the data matrix \mathbf{A} are powers of one another, this may lead to severe collinearity problems. The quickest way to access collinearity is to check the condition number of the scaled matrix \mathbf{A}^* . As per recommendation in [cite Seber and Lee], the data is not centred to see the effect on the raw data on LSE variance:

$$a_{i,j}^* = \frac{a_{i,j}}{\left\{ \sum_{i=1}^K a_{i,j}^2 \right\}^{1/2}},$$

where j is the column index. For the given data $\kappa(\mathbf{A}^*) =$ which indicates that some eigenvalues of the regression matrix $R_{aa} = (\mathbf{A}^*)^\top \mathbf{A}^*$ are close to zero. Indeed,

Some collinearity metrics for centred and scaled matrix \mathbf{A}_s^* , where $\mathbf{A} = \mathbf{1}, \mathbf{A}_s$ are summarised in Table 1. The level of collinearity is assessed by computing the coefficients of determination for individual columns

$$R_j^2 = 1 - \frac{(a_j^*)^\top H_j a_j^*}{\sum_i (a_j^* - \hat{a}_j^*)^2},$$

while the effect of collinearity on LSE variance is assessed via variance inflation coefficients

$$\text{VIF}_j = \frac{\text{Var}(\beta_j)}{\sigma^2} = \frac{\sigma^2}{1 - R_j^2}.$$

Evident collinearity indicates the need for regularisation of the estimation procedure in order to

Column	L	R	$L \times R$	$L \times L$	$R \times R$
R_j^2	0.9992	0.9988	0.9997	0.9992	0.9985
VIF_j	1256.04	862.29	2867.46	1212.49	665.06

reduce the dispersion of the parameter LSEs.

2 Regularised Least Squares

A common treatment for models that suffer from bad generalisation and collinearity is to introduce a constrained LS problem, where the constraint is normally posed on the unknown parameter vector:

$$\hat{\beta}^i = \arg \min_{\beta \in \mathbb{R}} \left\| \left(\bar{\theta}^i - X\beta^i \right) \right\|^2, \quad R(\beta^i) < \gamma. \quad (4)$$

where γ is a pre-specified parameter that defines the size of the constraint in the parameter space Ξ . Lagrangian formulation of the constrained problem is called regularised Least Squares (RLS),

$$\beta^i = \arg \min_{\beta \in \mathbb{R}} \left\{ \left\| \left(\bar{\theta}^i - X\beta^i \right) \right\|^2 + \lambda R(\beta^i) \right\} \quad (5)$$

where the regularisation coefficient λ is directly linked to γ in (4) CITE. Different types of the constraint function can be considered depending on the problem.

2.1 Tikhonov regularisation

Innovation regularisation constrains the 2-norm of the parameter vector

This is the only RLS formulation that has a closed form solution

$$\bar{\theta} = \mathbf{A}\beta, \quad (6)$$

This solution is referred to as ridge regression, because increasing λ leads to the

This shows that in Tikhonov regularisation λ quantifies the trade-off between the bias and the variance in the estimates. Small regularisation term leads to overfitting the model to the training data which may lead to poor

2.2 LASSO regularisation

While in the ridge regression LSEs asymptotically approach zero, none of the parameters can be zeroed-out explicitly if the model structure is overly detailed. Another formulation of RLS, called Least absolute shrinkage and selection operator (LASSO) regression, performs variable selection and the regularisation simultaneously. The optimisation problem is

In LASSO approach, regularisation

Remark 2.1: *Results of ridge estimation can be used as the initial point in convex optimisation while solving the LASSO regression.*

2.3 Selection of the regularisation coefficient

The important stage of solving RLS problem is selecting the regularisation parameter. The constraint γ in (4) is often selected arbitrary since the shape of the parameter space is unknown. As a result, finding λ relies on iterative schemes most of which do not guarantee convergence [CITE MANY]. For the lack of universal approach for selecting the optimal value of λ , the choice of the method remains application-specific.

In this report, the model structure is ass

3 Results

The results are obtained for the auxetic form data for 8 datasets using both types of regularisation. Ridge traces are shown for values of λ from 10^{-6} to 10^0 . While the RLS is usually used for the normalised (scaled and centred) data, results for the raw

4 Future work

- Regularised LS for the direct estimation of B from joint datasets. Long time-series and polynomial structure of the regressors may also lead to collinearity problems. It's not clear yet whether the data needs to be normalised separately.
- Confidence and covariance regions to demonstrate bias-variance trade off.
- Multiple methods for LASSO regression can be investigated.
-