A LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) FOR NONLINEAR SYSTEM IDENTIFICATION

Sunil L. Kukreja [*,1] **Johan Löfberg** [**] **Martin J. Brenner** [*]

[*] *NASA, Dryden Flight Research Center, Structural Dynamics Group, MailStop 4820 2A, Edwards, California 93523, USA*
[**] *Swiss Federal Institute of Technology, Automatic Control Lab, Physikstrasse 3, ETL I 22, CH-8092 Zürich, Switzerland*

Abstract: Identification of parametric nonlinear models involves estimating unknown parameters and detecting its underlying structure. Structure computation is concerned with selecting a subset of parameters to give a parsimonious description of the system which may afford greater insight into the functionality of the system or a simpler controller design. In this study, a least absolute shrinkage and selection operator (LASSO) technique is investigated for computing efficient model descriptions of nonlinear systems. The LASSO minimises the residual sum of squares by the addition of a ℓ_1 penalty term on the parameter vector of the traditional ℓ_2 minimisation problem. Its use for structure detection is a natural extension of this constrained minimisation approach to pseudolinear regression problems which produces some model parameters that are exactly zero and, therefore, yields a parsimonious system description. The performance of this LASSO structure detection method was evaluated by using it to estimate the structure of a nonlinear polynomial model. Applicability of the method to more complex systems such as those encountered in aerospace applications was shown by identifying a parsimonious system description of the F/A-18 Active Aeroelastic Wing using flight test data.

Keywords: System Identification, Nonlinear Systems, Structure Detection, Aeroelasticity

1. INTRODUCTION

Discrete-time nonlinear polynomials are often useful to describe the input-output behaviour of complex systems encountered in many control engineering, aerospace engineering and biological science applications. These polynomial mappings describe the dynamic relationship of a system by an expansion of the present output value in terms of present and past values of the input signal and past values of the output signal. These models are popularly known as polynomial NARMAX (Nonlinear AutoRegressive, Moving Average eXogenous) models, a special case of the so-called NARMAX model class (Leontaritis and Billings, 1985*a*; Leontaritis and Billings, 1985*b*).

Many systems are described by these polynomial models having only a few terms. However, even if the system order is known through some *a priori* knowledge, a full expansion of this model representation yields a large number of candidate terms which may be required to represent the system dynamics. Often many of these candidate terms are insignificant and, therefore, can be removed. Hence, the structure detection problem is that of selecting a subset of candidate terms that best predicts the output whilst maintaining an efficient system description.

[1] Supported by grants from the National Research Council (USA) and NASA, Dryden Flight Research Center (Grant No: NASA-NASW-99027)

The relevance of structure computation is, for example, controller design and study of aerospace vehicle dynamics. For control, a parsimonious system description is essential for many control strategies. In modelling the objective is often to gain insight into the function of the underlying system.

There are two fundamental approaches to the structure detection problem: (i) exhaustive search, where every possible subset of the full model is considered (see e.g. (Draper and Smith, 1981)), or (ii) parameter variance, where the covariance matrix, P_θ, based on input-output data and estimated residuals is used to assess parameter relevance (see e.g. (Ljung, 1999)). Both have problems. Exhaustive search requires a large number of computations and parameter variance estimates are often inaccurate when the number of candidate terms is large.

Recently, a bootstrap method has been proposed to solve the structure detection problem for over-parameterised models (Kukreja *et al.*, 2004). Although it has been demonstrated that the bootstrap is a useful tool for structure detection of NARMAX models, there is a limitation of the model complexity that can be studied with this technique. This limitation is a result of the large number of candidate terms, for a given model order, and the data length required to guarantee convergence. It was demonstrated that a necessary condition for bootstrap structure detection to yield accurate results is the number of data points needed for identification be *at least* 10 times the square of the initial number of candidate terms.

For many practical systems, collecting large data records may be financially and/or technically infeasible. Estimation techniques used for NARMAX identification all require an over-determined system of equations to solve for the unknown system parameters. Due to the large number of possible candidate terms and limited data records available for any practical identification problem, it may not be feasible to analyse highly complex systems with the bootstrap technique. Nonlinear aeroelastic dynamics of aircraft are highly complex processes likely involving a large number of candidate terms which may not be accurately characterised by current approaches.

We propose the application of a novel method for NARMAX model identification via a least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996). This approach permits identification of NARMAX models in situations where current methods cannot be applied. In this paper, we show that the LASSO yields good results for structure detection of an over-parameterised polynomial NARMAX model in the presence of additive output noise. Application of structure computation to aeroelastic modelling is presented using flight test data from the F/A-18 Active Aeroelastic Wing (AAW) (Pendleton *et al.*, 2000) and shown to yield a parsimonious model structure whilst maintaining a high percent fit to cross-validation data.

The organisation of this paper is as follows. In §2 we formulate the identification problem addressed here. LASSO and its application as a tool for structure computation is discussed in §3. Simulation results of LASSO's performance as a structure detection instrument are presented in §4 whilst in §5 application results to flight test data of the F/A-18 AAW are presented. Section 6 provides a discussion of our findings and §7 summarises the conclusions of our study.

2. PROBLEM STATEMENT

Consider the linear statistical model

$$z(n) = \sum_{j=1}^{p} \theta_j f(\varphi_j(n)) + e(n) \tag{1}$$

where $z(n)$ is the observed system output, θ_j is an unknown system parameter, $\varphi_j(n)$ is a regressor, $e(n)$ is an independent Gaussian variable with zero-mean and constant variance σ^2, f is a nonlinear mapping of the regressors, and n is a sample index point.

Let the regressors be described as a linear expansion of the observed system output, input and noise as

$$\varphi(n) = [1, z(n-1), \cdots, z(n-n_z), u(n), \cdots, \\ u(n-n_u), e(n-1), \cdots, e(n-n_e)]^T \tag{2}$$

where u is the input. For the special case where f is a nonlinear mapping of polynomial form it may include a variety of nonlinear terms, such as terms raised to an integer power, products of current and past inputs, past outputs, or cross-terms. In addition, the nonlinear mapping, f, can be described by a wide variety of nonlinear functions such as a sigmoid.

Identification of a NARMAX model consists of three stages: (1) model order selection, (2) structure detection and (3) parameter estimation. A brief summary of each stage of this process is discussed next.

2.1 Model Order Selection

The central problem in NARMAX identification is that of selecting the correct model order. Initially, there are an infinite number of candidate terms that could be considered. Establishing the model order limits the choice of terms to be considered. For polynomial NARMAX models, the system order may be defined as an ordered tuple

$$O \doteq [n_u \, n_z \, n_e \, l] \tag{3}$$

where n_u is the maximum lag on the input, n_z the maximum lag on the output, n_e the maximum lag on the error and l is the maximum nonlinearity order. Note that for non-polynomial NARMAX models, l may be simply replaced by a nonlinear mapping of some specified class. For simplicity, in the sequel, we assume nonlinearities that can be described by a polynomial expansion.

2.2 Structure Detection

The structure detection problem is that of selecting the subset of candidate terms that best describes the output. Therefore, the parametrisation of a system is still further reduced by determining which of the components are required. The maximum number of terms in a NARMAX model with n_z, n_u and n_e dynamic terms and lth order nonlinearity is:

$$p = \sum_{i=1}^{l} p_i + 1; p_i = \frac{p_{i-1}(n_u + n_z + n_e + i)}{i}, p_0 = 1 \quad (4)$$

As a result, the number of candidate terms becomes very large for even moderately complex models making structure detection difficult. We define the maximum number of terms, p, as the number of candidate terms to be initially considered for identification. Due to the excessive parameterisation (the curse of dimensionality), the structure detection problem often leads to computationally intractable combinatorial optimisation problems.

2.3 Parameter Estimation

The final step involves the estimation of the individual model parameters. Since a NARMAX model is linear in its parameters, standard least-squares minimisation techniques can be used:

$$\min_{\theta} \frac{1}{2} \left\| (\mathbf{Z} - \Phi\theta) \right\|_2^2 \quad (5)$$

where $\mathbf{Z} \in \mathbb{R}^{N \times 1}$ is a vector of outputs, $\Phi \in \mathbb{R}^{N \times p}$ is a matrix of regressors and $\theta \in \mathbb{R}^{p \times 1}$ is a vector of unknown system coefficients. The regression matrix is a function of the measured input-outputs and unmeasured noise, which makes this a pseudolinear regression problem since Φ is (partly) unknown and must be estimated along with the parameters. The noise is estimated as a sequence of prediction errors as, $\mathbb{R}^{N \times 1} \ni \hat{\varepsilon} = \mathbf{Z} - \hat{\mathbf{Z}}$ where $\hat{\mathbf{Z}} = \Phi\hat{\theta}$ is the predicted output and $\hat{\theta}$ the estimated parameter vector. As stated earlier, using least-squares it is difficult to estimate accurate parameter variance when the number of candidate terms is large. Therefore, a novel procedure which may enable structure selection of highly over-parameterised models is now considered.

3. LASSO

The least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) is least-squares like problem with the addition of a ℓ_1 penalty on the parameter vector as

$$\min_{\theta} \frac{1}{2} \left\| (\mathbf{Z} - \Phi\theta) \right\|_2^2 + \lambda \left\| \theta \right\|_1 \quad (6)$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm and $\|\cdot\|_1$ denotes the ℓ_1-norm.

The regularisation parameter $\mathbb{R} \ni \lambda = [\lambda_{min}, \ldots, \lambda_{max}]$ controls the trade-off between approximation error and sparseness. The LASSO shrinks the least-squares estimator (Eqn. 5) towards 0 and potentially sets $\theta_j = 0$ for some j. Consequently, LASSO behaves as a structure selection instrument.

3.1 Solution of LASSO

A solution to LASSO can be constructed in a quadratic programming framework (Chen *et al.*, 2001). With the introduction of slack variables the solution to this optimisation problem can be written as a simple bound constrained quadratic program (QP),

$$\min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \mathbf{M} \mathbf{x} + \mathbf{c}^T \mathbf{x} \text{ such that } x_k \geq 0, \text{and where,} \quad (7)$$

$$\mathbf{M} = \begin{bmatrix} \Phi^T\Phi & -\Phi^T\Phi \\ -\Phi^T\Phi & \Phi^T\Phi \end{bmatrix}, \mathbf{c} = \lambda \mathbf{1} - \begin{bmatrix} \Phi^T\mathbf{Z} \\ -\Phi^T\mathbf{Z} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \theta^+ \\ \theta^- \end{bmatrix}.$$

The model parameters are given by $\theta = \theta^+ - \theta^-$. The QP can be solved readily using standard optimisers (Mészáros, 1998). Thus, given a suitable regularisation parameter, the general structure computation problem can be solved. A method which enables the selection of an appropriate regularisation parameter is now considered.

3.2 Selection of Regularisation Parameter: λ

LASSO requires the determination of regularisation parameter, λ, for the penalty term (Eqn. 6) . To obtain λ, the method of cross-validation is used (Shao, 1993). This approach allows the prediction error

$$PE = E\left[\mathbf{Z} - \Phi\theta\right]^2 \quad (8)$$

to be estimated. The regularisation parameter, λ, is chosen to minimise this estimate.

3.3 Unique Optimum & Convergence of LASSO

For identification it is often assumed the excitation signal is persistently exciting which implies that $\Phi^T\Phi$ is positive definite. As a result the first term of Eqn. 6 is a strictly convex function. Since the second term is convex, it follows that the sum is strictly convex and a unique optimiser is guaranteed (Osborne *et al.*, 2000). Next, assume the optimal regularisation parameter, λ^*, is known. Since Eqn. 6 is strictly a convex optmisation problem the solution will converge to a unique global minimum (Osborne *et al.*, 2000). From parametric optimisation theory it is known that $PE(\lambda)$ is not necessarily a convex function (it is a piecewise quadratic function) (Grigoriadis and Ritter, 1969). Hence, for several choices of λ giving the same PE but different model structures can result. In the sequel, we investigate the validity of LASSO to select the correct model structure for a simulated nonlinear model.

4. SIMULATION EXAMPLE

The efficacy of LASSO as a tool for structure detection was assessed using Monte-Carlo simulations of a polynomial nonlinear system. In these simulations, a white input with uniform distribution was used. One thousand Monte-Carlo simulations were generated in which each input-output realisation was unique and had a unique Gaussian distributed, white, zero-mean, noise sequence added to the output. The output additive noise amplitude was increased in increments of 5 dB, from 20 to 0 dB signal-to-noise-ratio (SNR). Each input-output set consisted of 1,000 data points.

The regularisation parameter, λ, was determined by numerically minimising the cross-validation error across a discrete set of 1,000 logarithmically spaced λ values ($10^{\lambda_{min}} \leq \lambda \leq 10^{\lambda_{max}}$). The min-max regularisation parameter levels were set to $\lambda_{min} = -10$ and $\lambda_{max} = 1.5$. For cross-validation, the last 1/3 of each data set was used; 667 points for estimation and 333 for validation.

For each input-output realisation, the structure detection result was classified into one of three categories:

(1) Exact Model: A model which contains only true system terms,
(2) Over-modelled: A model with all its true system terms plus spurious parameters and
(3) Under-modelled: A model without all its true system terms. An under-modelled model may contain spurious terms as well.

We studied the nonlinear polynomial system:

$$z(n) = 0.4[u(n-1) + u(n-1)^2 + u(n-1)^3] + 0.8z(n-1) - 0.8e(n-1).$$

It was assumed that the system order is fully known, $O = [1,1,1,3]$, to assess the accuracy of LASSO to compute the correct structure when the model is mildly over-parameterised. This system is described by 3 lagged inputs, 1 lagged output, 1 lagged error and third order nonlinearity. A model of this order has 35 candidate terms, but the true system has only 5 true terms.

Fig. 1 presents the results of this study. The left panel

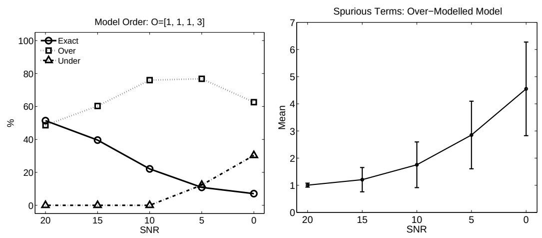

Fig. 1. Left: Selection rate of exact, over and under modelling. Right: Mean and STD of spurious term selection rate for over-modelling.

shows the LASSO method had a 0% rate of under-modelling for 20–10 dB SNR which then increased from 12.3%–30.3% for 5–0 dB SNR. The rate of over-modelling increased for 20–5 dB SNR, from 48.6%–76.8% then deceased at 0 dB SNR to 62.6% where under-modelling started to increase rapidly. The rate of selecting the exact model decreased across all SNR levels with a maximum of 51.4% at 20 dB and a minimum of 7.1% at 0 dB . The right panel illustrates the rate of selecting spurious terms when over-modelling occurred. This rate was low for all SNR levels with a minimum of 1.01 ± 0.0733 and a maximum of 4.55 ± 1.72. For this third order model with known model order, LASSO performed well for high SNR's since it did not drop any true terms. The performance of LASSO deteriorated when the SNR decreased. When LASSO selected an over-modelled model, the rate of spurious term selection was low for all SNR levels.

5. EXPERIMENTAL AIRCRAFT DATA

Lastly, the LASSO technique was assessed on experimental flight test data from the F/A-18 AAW project at NASA Dryden Flight Research Center. The data analysed for this study used collective aileron position input and structural accelerometer response output.

5.1 Procedures

Flight data was gathered during subsonic flutter clearance of the AAW. At each flight condition, the aircraft was subjected to multisine inputs corresponding to collective and differential aileron, collective and differential outboard leading edge flap, rudder, and collective stabilator excitations in the range of 3-35 Hz for 26 sec. This paper considers accelerometer data measured during the collective aileron sweeps at Mach 0.85 at 4,572 m (15,000 ft). The input collective aileron position was obtained as the average of four position transducer measurements from the right and left ailerons. The output was taken as the response of an accelerometer mounted near the wing leading edge just inside the wing fold. Data was sampled at 400 Hz. For analysis, the recorded flight test data was decimated by a factor of 2, resulting in a final sampling rate of 200 Hz.

For identification a model with fourth-order input-output and error dynamics and third-order nonlinearity, $O = [3,4,4,4]$, was used. A model with fourth-order dynamics was selected because it has been observed that aeroelastic structures are well defined by a fourth-order linear time-invariant (LTI) system (Smith, 1995). The nonlinearity order was chosen as third-order because models of higher nonlinear order can often be decomposed to second or third-order. This gave a full model description with 560 candidate terms.

The system was identified as outlined in §4. For estimation, $N_e = 5,200$ points were used from accelerometer response measurements on the right wing. For cross-validation, $N_v = 5,200$ points were used from data collected at a similar location on the left wing. In both the estimation and cross-validation sets, the input was the same collective aileron position. The min-max regularisation parameter levels were set to $\lambda_{min} = -10$ and $\lambda_{max} = 1.0$ with a discretisation grid of 1,000 logarithmically spaced λ's.

5.2 Results

The results of identifying the AAW data are presented. Fig. 2 shows the input-output trial used for this analysis. The data represents collective aileron position

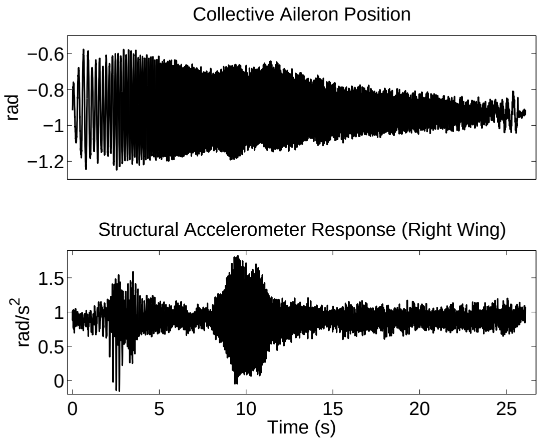

Fig. 2. Upper panel: Collective aileron position. Lower panel: Structural accelerometer response.

sequence and structural accelerometer response (right wing) used to compute the system structure.

Eqn. 9 depicts the model structure computed by the LASSO method

$$\begin{aligned} z(n) &= \hat{\theta}_0 + \hat{\theta}_1 u(n-1) + \hat{\theta}_2 u(n-2) + \hat{\theta}_3 u(n-4) \\ &+ \hat{\theta}_4 u^2(n-1) + \hat{\theta}_5 u^2(n-2) + \hat{\theta}_6 u^2(n-4) \\ &+ \hat{\theta}_7 z(n-1) + \hat{\theta}_8 z(n-4) + \hat{\theta}_9 u^2(n-1) z(n-4) \\ &+ \hat{\theta}_{10} u^2(n-2) z(n-1) + \hat{\theta}_{11} u^2(n-4) z(n-4) \\ &+ \hat{\theta}_{12} z^3(n-1) + \hat{\theta}_{13} z^3(n-4) + \hat{\theta}_{14} \hat{\varepsilon}(n-1) \\ &+ \hat{\theta}_{15} \hat{\varepsilon}(n-4) + \hat{\theta}_{16} u^2(n-1) \hat{\varepsilon}(n-4) \\ &+ \hat{\theta}_{17} u^2(n-2) \hat{\varepsilon}(n-1) + \hat{\theta}_{18} u^2(n-4) \hat{\varepsilon}(n-4) \\ &+ \hat{\theta}_{19} z^2(n-1) \hat{\varepsilon}(n-1) + \hat{\theta}_{20} z(n-1) \hat{\varepsilon}^2(n-1) \\ &+ \hat{\theta}_{21} \hat{\varepsilon}^3(n-1) + \hat{\theta}_{22} z^2(n-4) \hat{\varepsilon}(n-4) \\ &+ \hat{\theta}_{23} z(n-4) \hat{\varepsilon}^2(n-4) + \hat{\theta}_{24} \hat{\varepsilon}^3(n-4). \end{aligned} \tag{9}$$

The computed model structure is represented by a combination of linear and nonlinear, lagged input-output terms and contains 25 terms. Hence, the LASSO technique successfully produced a parsimonious model description from the full set of 560 candidate terms.

Fig.3 shows the predicted output for a cross-validation data set for the identified structure (Eqn. 9). The upper panel displays the full 26s time history of the accelerometer response recorded on the left wing. The

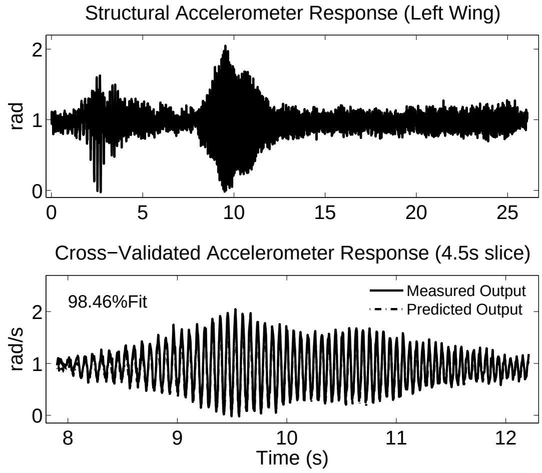

Fig. 3. Upper panel: Full time history of structural accelerometer response. Lower panel: Predicted accelerometer response of left wing superimposed on top of measured velocity output.

lower panel displays a 4.5 second slice of the predicted output superimposed on top of the measured output. The predicted output accounts for over 98% of the measured outputs variance. The result demonstrates that the computed model structure is capable of reproducing the measured output with high accuracy.

6. DISCUSSION

6.1 Simulations

The LASSO approach to structure detection yielded good results for the nonlinear polynomial model considered. It had a high rate of selecting the exact structure for high SNR levels. For lower SNRs, over-modelling dominated the structure computation procedure except at 0 dB at which the rate of under-modelling started to increase quickly. For all SNR levels when the computed model was over-modelled, the rate of selecting spurious terms was low.

The results for this case study were obtained using only 667 data points for estimation and 333 for validation. Often, there is more data available in many system identification applications and, therefore, the exact selection rate should improve given more data for identification.

6.2 Experimental Aircraft Data

Experimental results demonstrate that LASSO may be a useful tool for structure computation of dynamic aircraft data. LASSO successfully reduced the large number of regressors posed to aircraft aeroelastic data yielding a parsimonious model structure. Additionally, this parsimonious structure was capable of predicting a large portion of the observed cross-validation data, collected on an adjacent wing and with a different sensor. This suggests that the identified structure and parameters explain the data well. Using percent fit alone as an indicator of model goodness may lead to incorrect interpretations of model validity. However, in many cases, for nonlinear models this may be the only indicator that is readily available.

For this study, only a polynomial map was used as a basis function to explain the nonlinear behaviour of the F/A-18 AAW data. Clearly, different basis functions should be investigated to determine if another basis can produce accurate model predictions with reduced or comparable complexity. Moreover, further studies are necessary to evaluate whether the model structure is invariant under different operating conditions, such as Mach number and altitude, and for model parameterisations.

6.3 Improvements and Future Work

There may be more efficient ways to address the solution of Eqn.6 for different λ. The LASSO optimisation problem can be viewed as special case of a parameterised QP. It is known that the solution $\theta(\lambda)$ is a piecewise function (Grigoriadis and Ritter, 1969) and there are reasonably efficient ways to construct this function for low order dimension on the parametric variable λ (Kvasnica *et al.*, 2004). Since λ is scalar in our case, a parametric approach might be tractable. If the piecewise function $\theta(\lambda)$ has a compact description, it could be more computationally efficient than a brute-force gridding.

It should be acknowledged that the overall problem with the two stages, (i) computing a finite set of optimal θ for varying λ and (ii) optimising λ in a second phase to minimise the cross-validation criteria, can be interpreted as an ad-hoc approach to solve a bi-level optimisation problem. These problems are notoriously difficult to solve exactly. Nevertheless, improved computation of sub-optimal solutions may be possible by exploring more advanced approaches to address the bi-level optimisation problem

The optimisation criteria in the LASSO setup is motivated by the well known fact that a $\|\cdot\|_1$ penalty appended to a quadratic objective tends to yield a sparse solution. However, it is only a heuristic for addressing the underlying problem: achieving few non-zero parameters. An alternative way to address this, in an optimisation framework, is to use combinatorial optimisation. Solving the regression problem (Eqn. 5) with a bound on the number of non-zero parameters may be achieved in a straightforward manner using mixed binary quadratic programming. Instead of performing the cross-validation optimisation problem over a bank of solutions, computed using different λ, one could compute solutions for a different number of non-zero parameters and use these solutions in the cross-validation phase.

7. CONCLUSIONS

LASSO is a novel approach for detecting the structure of over-parameterised nonlinear models in situations where other methods may be inadequate. The main point here is that the LASSO technique is clearly amenable to the study of a wide range of nonlinear systems. These results may have practical significance in the analysis of aircraft dynamics during envelope expansion and could lead to more efficient control strategies. In addition, this technique could allow greater insight into the functionality of various systems dynamics, by providing a quantitative model which is easily interpretable.

REFERENCES

Chen, S.S., D.L. Donoho and M.A. Saunders (2001). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **43**(1), 129–159.

Draper, N.R. and H. Smith (1981). *Applied Regression Analysis*. second ed.. John Wiley & Sons. New York.

Grigoriadis, M.D. and K. Ritter (1969). A parametric method for semidefinite quadratic programs. *SIAM J. Control* **7**(4), 559–577.

Kukreja, S.L., H.L. Galiana and R.E. Kearney (2004). A bootstrap method for structure detection of NARMAX models. *Int. J. Cont.* **77**(2), 132–143.

Kvasnica, M., P. Grieder, M. Baotić and M. Morari (2004). Multi Parametric Toolbox (MPT). In: *Hybrid Systems: Computation and Control*. Vol. 2993 of *Lecture Notes in Computer Science*. Springer Verlag. Philadelphia. pp. 448–462.

Leontaritis, I.J. and S.A. Billings (1985*a*). Input-output parametric models for non-linear systems part I: deterministic non-linear systems. *Int. J. Control* **41**(2), 303–328.

Leontaritis, I.J. and S.A. Billings (1985*b*). Input-output parametric models for non-linear systems part II: stochastic non-linear systems. *Int. J. Control* **41**(2), 329–344.

Ljung, L. (1999). *System Identification: Theory for the User*. second ed.. Prentice Hall, Inc.. Englewood Cliffs, New Jersey.

Mészáros, C. (1998). The BPMPD interior point solver for convex quadratic problems. Technical Report 98–8. Hungarian Acad.of Sci.. Budapest.

Osborne, M.R., B. Presnell and B.A. Turlach (2000). On the lasso and its dual. *J. Comp. Graph. Stat.* **9**(2), 319–337.

Pendleton, E.W., D. Bessette, P.B. Field, G.D. Miller and K.E. Griffin (2000). Active aeroelastic wing flight research program: Technical program and model analytical development. *AIAA J. Aircraft* **37**(4), 554–561.

Shao, J. (1993). Linear model selection by cross-validation. *J. American Stat. Assoc.* **88**, 486–494.

Smith, R. S. (1995). Model validation for robust control: An experimental process control application. *Automatica* **31**(11), 1637–1647.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B* **58**(1), 267–288.