

1 Data_driven technique to discover optimal model of \ddot{F}

To estimate \ddot{f} , a data_driven technique is employed. We consider \ddot{F} is linear in terms of θ , α , $\dot{\theta}$, and $\dot{\alpha}$ as equation 1 presents.

$$\ddot{F} = c_1\theta + c_2\alpha + c_3\dot{\alpha} + c_4\dot{\theta} \quad (1)$$

where c_1 , c_2 , c_3 , and c_4 are constant coefficients. From experiment, only the time incremental record of angles θ and α are available, and F is known by $\theta + 1.039\alpha$. To estimate the derivatives and filter the measurement noises of θ and α , the algebraic operation in [?] is applied. We assume the dynamic of system 1 is sparse in the states and employ the least absolute shrinkage and selection operator (LASSO) to estimate the sparse terms through minimizing the cost function J_λ in 2.

$$J_\lambda(\mathbf{c}) = \|\mathbf{e}\|_2^2 + \lambda \|\mathbf{c}\|_1 \quad (2)$$

Let us introduce \mathbf{c} as the vector of coefficients and \mathbf{e} accounts for the estimation error. $\lambda > 0$ is a preselected positive number known as the sparse regulator. It is common to use cross-validation techniques from machine learning to determine the sparse regulation parameter λ . The value of λ in an finite interval is sampled and total data is divided into train and test dataset. The cross-validation mean square error over the test dataset is computed. The λ value which minimizes the cross-validation error is selected. The corresponding model is chosen as an optimal model with a proper balance of complexity and accuracy.

In order to develop robust sparse regression, we propose to combine the sparse regression with bootstrapping. We consider K bootstrap sample vectors containing L elements of the original data points. For each bootstrap sample vector, the sparse regression is applied to identify the model. Finally, the parameters of the model is the average of the K estimated coefficient vectors $\hat{\mathbf{c}}_l$.

$$\tilde{\mathbf{c}} = \frac{1}{K} \sum_{l=1}^K \hat{\mathbf{c}}_l \quad (3)$$

The standard deviation of estimated coefficient vectors $\hat{\mathbf{c}}_l$ is computed to study the variation of parameter estimation. We record data for 4 s with the time increment 0.002 s. The first half of the data is kept for training and we use the rest for test data. For bootstrapping, 30 samples with the length of 30% are selected. The algorithm estimates

$$\ddot{F} = (-44.2532 \pm 0.0361)\alpha \quad (4)$$