4th International Conference on Mathematical Modeling in Physical Sciences (IC-MSquare2015) IOP Publishing
Journal of Physics: Conference Series **633** (2015) 012057 doi:10.1088/1742-6596/633/1/012057

A multi-mode cantilever singular point detection using adaptive hypothesis testing

Jakub Dokoupil and Pavel Václavek

Central European Institute of Technology, Brno University of Technology, 616 00 Brno, Czech Republic

E-mail: jakub.dokoupil@ceitec.vutbr.cz, pavel.vaclavek@ceitec.vutbr.cz

Abstract. Fundamental analysis of a multi-mode model of the atomic force microscope cantilever shows that at some points; called here singular points, the mode is vanished. Consequently, the order of the input/output behavior is reduced. The singular points can be detected comparing possible candidates on the best model order. The detection is then naturally performed by applying the Bayesian model comparison. Since the exact position of the singular points is not available a priori, an explicit model of updating the probability of tested hypotheses in time is built. More specifically, a mechanism of suppressing absolute information is suggested based on the Bayesian decision problem where the Kullback-Leibler divergence is used.

1. Introduction

The problem of singular point detection of the atomic force microscope cantilever [1] can be viewed as the adaptive testing of hypotheses, each one matched to a certain model order. As the system model is consider to be stochastic, these hypotheses naturally represent the realizations of the mass function. To compensate the lack of knowledge about the true singular point positions, the idea of the posterior mass function flattening can be employed. The flattening is generally performed by exponential or linear forgetting [2]. In the present paper, we extend the approach of the exponential forgetting scheme. We attempt to give a stabilized forgetting strategy where an explicit model of updating the hypotheses in time is built. More specifically, the resulting formula is based on solving the statistical decision problem [3], [4], where the Kullback-Leibler divergence [5] is used to measure a distance between two mass functions.

Dealing with the problem of which from competing models is more suitable to explain the data generating process a subjective measure of confidence for each model is evaluated to choose the most probable one. The known solution satisfying the theorems of probability theory enables us to compare models with noninformative priors [6], [7] only over parameters which are common to all models. Hence, the given solution for the comparison of nonnested models produced by the Kalman based algorithms could not be put into practice. For this reason, the existing concept is reformulated preserving its Bayesian principle to be more beneficial in the case where on-line processing is needed.

2. Adaptive model comparison

Consider that the directly manipulated input u_k and the indirectly effected output y_k are both observed on the system at every discrete instant of time k. Let the sequence of input-output data pairs from time k_1 up to the time k be denoted by $D_{k_1}^k = \{u_i, y_i\}_{i=k_1}^k$. The j-mode cantilever model [1] can be described in its input/output form as follows:

$$\lfloor j\theta_{k+1} = \lfloor j\theta_k, \tag{1}$$

$$y_k = \lfloor j h_k' \, \lfloor j\theta_k + e_k, \quad e_k \sim \mathcal{N}\big(0, \lfloor j r_k\big)\big|_{\lfloor j r_{k+1} = \lfloor j r_k}, \tag{2}$$

where $\lfloor j h_k \in \mathbb{R}^{2j}$ represents the finite-dimensional regression vector containing observations $\{u_i, y_i\}_{i=k-j}^{k-1}$, h_k' stands for the transposition of h_k, $\lfloor j\theta_k \in \mathbb{R}^{2j}$ is the vector of inner random variables. The unmeasurable variable e_k is presumed to be a discrete white noise, normally distributed with a zero mean and an unknown variance $\lfloor j r_k$.

Solving the problem when more than one model order is considered to be possible, the selection of the best one can be performed by applying the Bayesian model comparison. Suppose that we have the finite set of different competing model orders $\{\lfloor j\mathcal{M}_k\}_{j=1}^n$, which all seek to explain the system under study. To assess the degree of support for a particular model order $\lfloor j\mathcal{M}_k$ corresponding to the parameter set $\lfloor j\Theta_k = \{\lfloor j\theta_k, \lfloor j r_k\}$, the posterior model probability $f(\lfloor j\mathcal{M}_k | D_{1-n}^k)$ shall to be evaluated. The Bayes' rule gives

$$f\left(\lfloor j\mathcal{M}_k \,\middle|\, D_{1-n}^k\right) \propto f\left(D_k^k \,\middle|\, D_{1-n}^{k-1}, \lfloor j\mathcal{M}_k\right) f\left(\lfloor j\mathcal{M}_k \,\middle|\, D_{1-n}^{k-1}\right). \tag{3}$$

where $\propto$ stands for equal up to the normalization. Since the singular points are not known beforehand, an explicit model of order variations is not available. Therefore, we develop a forgetting scheme for the Bayesian model comparison that is linked to the following common idea. The probability of the realization $\lfloor j\mathcal{M}_k$ given data D_{1-n}^k is described by the posterior probability $f(\lfloor j\mathcal{M}_k | D_{1-n}^k)$. The problem is to determine the predictive probability $f(\lfloor j\mathcal{M}_{k+1} | D_{1-n}^k)$ of the realization $\lfloor j\mathcal{M}_{k+1}$ at the next time instant given the same data D_{1-n}^k. Let us introduce a prior information about the model orders through alternative mass functions. The problem then reads: given two alternatives on the mass function $f(\mathcal{M}_{k+1} | D_{1-n}^k)$ ($\mathcal{M}_{k+1}$ is the discrete random variable), namely

- the posterior mass function $f_0(\mathcal{M}_{k+1}) = f(\mathcal{M}_k | D_{1-n}^k)$ in the case of no order changes, and
- the alternative mass function $f_1(\mathcal{M}_{k+1}) \propto 1$ describing the maximal uncertainty about $\mathcal{M}_{k+1}$,

decide on an optimal mass function $\hat{f}(\mathcal{M}_{k+1})$. The task of choosing $\hat{f}(\mathcal{M}_{k+1})$ given $f_0(\mathcal{M}_{k+1})$ and $f_1(\mathcal{M}_{k+1})$ is formulated as a Bayesian decision problem. In particular, the probabilities λ and $1-\lambda$ are assigned to the alternatives $f_0(\mathcal{M}_{k+1})$ and $f_1(\mathcal{M}_{k+1})$ and the best decision $\hat{f}(\mathcal{M}_{k+1})$ is defined as the mass function minimizing expectation of a distance between the true mass function $f(\mathcal{M}_{k+1})$ and the unknown mass function $f_A(\mathcal{M}_{k+1})$ restricted to the parameter space $f_A^*(\mathcal{M}_{k+1}) \equiv \{f_0(\mathcal{M}_{k+1}), f_1(\mathcal{M}_{k+1})\}$. In order to evaluate a proper distance between $f\big(\mathcal{M}_{k+1}\big)$ and $f_A\big(\mathcal{M}_{k+1}\big)$, the Kullback-Leibler divergence is used

$$\mathcal{D}\big(f\big(\mathcal{M}_{k+1}\big) \| f_A\big(\mathcal{M}_{k+1}\big)\big) = \sum_{j=1}^{n} f\big(\lfloor j\mathcal{M}_{k+1}\big) \ln\left(\frac{f\big(\lfloor j\mathcal{M}_{k+1}\big)}{f_A\big(\lfloor j\mathcal{M}_{k+1}\big)}\right). \tag{4}$$

Recall that ${}^{\lfloor j}\mathcal{M}_{k+1}$ is the realization of the random variable $\mathcal{M}_{k+1}$. Taking the expectation $\mathcal{E}[\cdot]$ of (4) results in

$$\begin{aligned}&\mathcal{E}\big[\mathcal{D}\big(f\big(\mathcal{M}_{k+1}\big)\|f_A\big(\mathcal{M}_{k+1}\big)\big)\big] \qquad (5)\\ =&\lambda\sum_{j=1}^{n} f\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)\ln\left(\frac{f\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)}{f_0\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)}\right)+\big(1-\lambda\big)\sum_{j=1}^{n} f\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)\ln\left(\frac{f\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)}{f_1\big({}^{\lfloor j}\mathcal{M}_{k+1}\big)}\right)\\ =&\mathcal{D}\left(f\big(\mathcal{M}_{k+1}\big)\middle\|\frac{f_0\big(\mathcal{M}_{k+1}\big)^{\lambda}f_1\big(\mathcal{M}_{k+1}\big)^{(1-\lambda)}}{\sum_{l=1}^{n} f_0\big({}^{\lfloor l}\mathcal{M}_{k+1}\big)^{\lambda}f_1\big({}^{\lfloor l}\mathcal{M}_{k+1}\big)^{(1-\lambda)}}\right)-\ln\left(\sum_{l=1}^{n} f_0\big({}^{\lfloor l}\mathcal{M}_{k+1}\big)^{\lambda}f_1\big({}^{\lfloor l}\mathcal{M}_{k+1}\big)^{(1-\lambda)}\right).\end{aligned}$$

The optimal prediction $\hat{f}(\mathcal{M}_{k+1})$ is found by solving the following decision problem:

$$\hat{f}(\mathcal{M}_{k+1})=\arg\min_{f(\mathcal{M}_{k+1})}\mathcal{E}\big[\mathcal{D}\big(f\big(\mathcal{M}_{k+1}\big)\|f_A\big(\mathcal{M}_{k+1}\big)\big)\big]. \qquad (6)$$

The resulting form of (5) and the properties of the Kullback-Leibler divergence imply that for individual realizations generated by $\hat{f}(\mathcal{M}_{k+1})$ hold

$$f\left({}^{\lfloor j}\mathcal{M}_{k+1}\middle|D_{1-n}^{k}\right)=f^{\lambda}\left({}^{\lfloor j}\mathcal{M}_{k}\middle|D_{1-n}^{k}\right)/\sum_{l=1}^{n} f^{\lambda}\left({}^{\lfloor l}\mathcal{M}_{k}\middle|D_{1-n}^{k}\right). \qquad (7)$$

The probability $\lambda\in(0,1)$ can be viewed more prosaically as the forgetting factor with a lower value producing a higher degradation of less relevant information. At this point, equation (3) can be rewritten according to

$$f\left({}^{\lfloor j}\mathcal{M}_{k}\middle|D_{1-n}^{k}\right)\propto\int_{\Theta^*} f\left(y_k\middle|{}^{\lfloor j}h_k,{}^{\lfloor j}\Theta_k\right)f\left({}^{\lfloor j}\Theta_k\middle|D_{1-n}^{k-1}\right)d\,{}^{\lfloor j}\Theta_k f^{\lambda}\left({}^{\lfloor j}\mathcal{M}_{k-1}\middle|D_{1-n}^{k-1}\right), \qquad (8)$$

where Θ^* is the space of ${}^{\lfloor j}\Theta_k$. By invoking the system (1), (2), it allows us to algebraically quantify the functional form (8) as follows:

$$f\big({}^{\lfloor j}\mathcal{M}_{k}\big|D_{1-n}^{k}\big)=\frac{{}^{\lfloor j}c_k\,{}^{\lfloor j}\Sigma_{k-1}^{{}^{\lfloor j}\gamma_{k-1}/2}\,{}^{\lfloor j}\hat{r}_{y;k}^{-1/2}\,{}^{\lfloor j}\Sigma_{k}^{-\,{}^{\lfloor j}\gamma_{k}/2}f^{\lambda}\left({}^{\lfloor j}\mathcal{M}_{k-1}\big|D_{1-n}^{k-1}\right)}{\sum_{l=1}^{n}{}^{\lfloor l}c_k\,{}^{\lfloor l}\Sigma_{k-1}^{{}^{\lfloor l}\gamma_{k-1}/2}\,{}^{\lfloor l}\hat{r}_{y;k}^{-1/2}\,{}^{\lfloor l}\Sigma_{k}^{-\,{}^{\lfloor l}\gamma_{k}/2}f^{\lambda}\left({}^{\lfloor l}\mathcal{M}_{k-1}\big|D_{1-n}^{k-1}\right)}, \qquad (9)$$

where the following quantities have to be calculated for each model order:

$$\begin{aligned}{}^{\lfloor j}\hat{r}_{y;k}&\equiv 1+{}^{\lfloor j}h_k'\,{}^{\lfloor j}P_{k-1}\,{}^{\lfloor j}h_k, && (10)\\ {}^{\lfloor j}\hat{e}_k&\equiv y_k-{}^{\lfloor j}h_k'\,{}^{\lfloor j}\hat{\theta}_{k-1}, && (11)\\ {}^{\lfloor j}K_k&\equiv{}^{\lfloor j}P_{k-1}\,{}^{\lfloor j}h_k\,{}^{\lfloor j}\hat{r}_{y;k}^{-1}, && (12)\\ {}^{\lfloor j}\hat{\theta}_k&={}^{\lfloor j}\hat{\theta}_{k-1}+{}^{\lfloor j}K_k\,{}^{\lfloor j}\hat{e}_k, && (13)\\ {}^{\lfloor j}P_k&=(I-{}^{\lfloor j}K_k\,{}^{\lfloor j}h_k')\,{}^{\lfloor j}P_{k-1}, && (14)\\ {}^{\lfloor j}\Sigma_k&={}^{\lfloor j}\Sigma_{k-1}+{}^{\lfloor j}\hat{e}_k^2\,{}^{\lfloor j}\hat{r}_{y;k}^{-1}, && (15)\\ {}^{\lfloor j}\gamma_k&={}^{\lfloor j}\gamma_{k-1}+1 && (16)\end{aligned}$$

and the auxiliary parameter ${}^{\lfloor j}c_k$ is defined as the ratio of two gamma functions ${}^{\lfloor j}c_k=\Gamma({}^{\lfloor j}\gamma_k/2)/\Gamma\left({}^{\lfloor j}\gamma_{k-1}/2\right)$. A detailed discussion about the particular quantities above and their prior settings can be found in [8].

From (9) it can be inferred that the Bayesian model comparison favors those models for which the addition of new observations will cause an increase in accuracy. The measure of the jth model order accuracy in terms of the model error rate in the output predictions is reflected by the sum of squared errors $\sum_{\kappa=j-n+1}^{k}(y_\kappa - {\lfloor j}h'_\kappa\, {\lfloor j}\hat{\theta}_k)^2$. These errors form the part of ${\lfloor j}\Sigma_k$ with a lower value indicating a higher model accuracy. The model comparison also contains reward for coherency between the data and the prior information (via $({\lfloor j}\hat{\theta}_k - {\lfloor j}\hat{\theta}_{j-n})'({\lfloor j}P^{-1}_{h;j-n})({\lfloor j}\hat{\theta}_k - {\lfloor j}\hat{\theta}_{j-n})$ appearing in ${\lfloor j}\Sigma_k$) and parsimony.

Simulations produce excellent results and point out to the validity of our solution. However, due to space limitation, the simulation studies are not presented here.

3. Conclusion

In this paper a novel approach for detection of singular points of the multi-mode cantilever has been developed based on the Bayesian inference principle. More specifically, the algorithm which is capable of deciding which of the model orders is the most probable one was derived. To detect the singular points, which position is unknown, stabilized forgetting strategy that is close to the well-known technique of exponential forgetting was embed into the mechanism of time-evolution of the particular hypotheses. The algorithm generally provides a solution in the case where data are generated by the Gaussian probability density function through the unknown time-varying system parameters.

Acknowledgements

The first author was supported by the project CZ.1.07/2.3.00/30.0039 of Brno University of Technology. The work of the second author was realized at Central European Institute of Technology (CEITEC), with research infrastructure supported by the Project CZ.1.05/1.1.00/02.0068 financed from European Regional Development Fund. This support is gratefully acknowledged.

References

[1] Salapaka V M Bergh S H Lai J Majumdar A and McFarland E 1997 Multi-mode noise analysis of cantilevers for scanning probe microscopy *J. Appl. Phys.* **81** pp 2480-87
[2] Kulhavý R and Kraus J F 1996 On duality of regularized exponential and linear forgetting *Automatica* **32** pp 1403-15
[3] DeGroot H M 2004 *Optimal statistical decision* (Wiley-Interscience)
[4] Liese F and Miescke J K 2008 *Statistical decision theory: Estimation, testing, and selection* (Springer)
[5] Kullback S and Leibler R 1951 *On information and sufficiency* (Annals of Mathematical Statistics vol 22) pp 79-87
[6] Koop G 2003 *Bayesian econometrics* (Wiley-Interscience)
[7] Geweke J 2005 *Contemporary Bayesian econometrics and statistics* (Wiley-Interscience)
[8] Kárný M Bohm J Guy T Jirsa L Nagy I Nedoma P and Tesar L 2006 *Optimized Bayesian dynamic advising: Theory and algorithms* (Springer Verlag)