An Identification Approach to Dynamic Errors-in-variables Systems with a Preliminary Clustering of Observations

Levente Hunyadi

Errors-in-variables models are statistical models in which not only dependent but also independent variables are observed with error, i.e. they exhibit a symmetrical model structure. The application field for these models is diverse including computer vision, image reconstruction, speech and audio processing, signal processing, modal and spectral analysis, system identification and astronomy [1].

One particular application of errors-in-variables models is to parameter estimation of discrete-time dynamic linear systems. In this configuration, only the noise-corrupted input and output signals are observable, the original noise-free signals are not. As far as the additive noise is concerned, in most cases, a white noise model is assumed, which corresponds to noise due to measurement error. Given this system model, the goal of system identification is to derive process as well as noise parameters using the observable noise-contaminated input and output signals.

Provided that the ratio of input and output noise variances is a priori known, the task of deriving process and noise parameters is a classical system identification problem. In contrast, a situation where no such information is available is recognized as a more difficult one. In fact, it turns out that under the assumption of Gaussian distributed input and noise signals, the system is not identifiable by second order properties only (i.e. the autocorrelation function in time domain or the power spectrum in frequency domain). In other words, identification results in many indistinguishable solutions, so that for the identification to produce a unique result, additional restrictions are necessary [2].

One possible restriction is to assume that observations can be clustered into two sets which expose different characteristics. Provided that such a clustering is possible, identification can be broken down into parameter estimation over two separate sets of data. As a result, the estimation procedure produces two sets of process parameter estimates, the distance of which can be measured using some distance metrics. Consequently, varying the ratio of input and output noise variances, and deriving process parameter estimates for each case, minimizing the distance metrics yields an estimate for the noise variance ratio, resolving the identifiability issue.

The feasibility of the approach is demonstrated by a simulation example. Clustering is performed by principal component analysis [3] and estimation is based on a generalized Koopmans-Levin method [4]. The performance of the approach is also compared to the Cramér–Rao statistical lower bound [5], which gives the theoretical value for the best possible estimate. The algorithmic aspects of the simulation environment are implemented in MatLab with a visual front-end realized in C#.

References

[1] Sabine Van Huffel and Philippe Lemmerling, editors. TLS and EIV modeling. *Kluwer*, 2002.

[2] Torsten Söderström. Errors-in-variables methods in system identification. *Automatica*, 43:939–958, 2007.

[3] Chris Ding and Xiaofeng He. K-means clustering via principal component analysis. In *Proceedings of the 21st Int. Conf. on Machine Learning*, page 29, New York, NY, USA, 2004. ACM.

[4] István Vajk. Identification methods in a unified framework. *Automatica*, 41(8):1385–1393, 2005.

[5] Erlendur Karlsson, Torsten Söderström, and Petre Stoica. The Cramér-Rao lower bound for noisy input-output systems.*Signal Processing*, 80(11):2421–2447, 2000.