A simplified form of the bias-eliminating least squares method for errors-in-variables identification

Mei Hong, Torsten Söderström, Wei Xing Zheng

Department of Information Technology
Uppsala University
Box 337, SE-751 05 Uppsala, Sweden

Technical report 2006-040
August 2006
ISSN 1404-3203

A simplified form of the bias-eliminating least squares method for errors-in-variables identification

Mei Hong, Torsten Söderström
Division of Systems and Control,
Department of Information Technology, Uppsala University,
P O Box 337, SE-75105 Uppsala, Sweden

Wei Xing Zheng
School of QMMS, University of Western Sydney,
Penrith South DC NSW 1797, Australia

August 29, 2006

Abstract

This report considers the bias-eliminating least squares (BELS) method for identifying the errors-in-variables systems with white input noise and colored output noise. A simplified form of the BELS algorithm is proposed which is proved to be equivalent to the existing one. The new relation is a form of linear IV equations which will not only reduce the computational load but also simplify the analysis of the properties of the BELS estimates.

1 Introduction

Consider the noise-free input and output processes $u_0(t)$ and $y_0(t)$ which are linked by a linear dynamic system

$$A(q^{-1})\, y_0(t) = B(q^{-1})\, u_0(t), \tag{1.1}$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \ldots + a_{n_a}\, q^{-n_a} \\ B(q^{-1}) &= b_0 + b_1\, q^{-1} + \ldots + b_{n_b}\, q^{-n_b} \end{aligned} \tag{1.2}$$

are polynomials in the backward shift operator q^{-1}.

For errors-in-variables (EIV) systems, the input and the output are measured with additive noise $\tilde{u}(t)$ and $\tilde{y}(t)$, respectively. The available signals are of the form:

$$\begin{aligned} u(t) &= u_0(t) + \tilde{u}(t) \\ y(t) &= y_0(t) + \tilde{y}(t). \end{aligned} \tag{1.3}$$

We make the following general assumptions:

A1. The dynamic system (1.1) is asymptotically stable, *i.e.* $A(z)$ has all zeros outside the unit circle.

A2. All the system modes are observable and controllable, *i.e.* $A(z)$ and $B(z)$ have no common factors.

A3. The polynomial degrees n_a and n_b are a priori known.

A4. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually independent zero mean noise, and both independent of $u_0(t)$.

A5. The input noise $\tilde{u}(t)$ is white and the output noise is colored.

A6. The true input $u_0(t)$ is a zero-mean stationary ergodic random signal, that is persistently exciting of enough order.

The situation of the assumption **A5** is a fairly realistic one. In practice, the output noise $\tilde{y}(t)$ should model not only measurement and/or sensor noise but also effects of process disturbances. This situation does not apply for the input noise $\tilde{u}(t)$ normally.

The aim of identifying the EIV system is to consistently estimate the system parameter vector $\theta = (a_1 \dots a_{n_a} \quad b_0 \dots b_{n_b})^T$ from the measured noisy data $\{u(t), y(t)\}_{t=1}^N$. It is recognized as a difficult identification problem. During the past decades, many different solutions have been presented, see, for example, [3], [1], [6], etc. Within them the bias-eliminating least squares (BELS) algorithm is one of the more interesting approaches, because it usually gives quite good estimation accuracy but with a modest computational cost. The BELS methods (without pre-filter) were first proposed in [7] for the case of EIV systems with white input and white output noise. The algorithm was further extended to deal with cases with white input noise and colored output noise in [8]. The focus of this report lies on derivation and interpretations of a simplified BELS estimator, which has the advantage from the computational point of view and will benefit the theoretical analysis of the BELS method.

2 Notations

The following additional notations are introduced for convenience. The regressor vector is defined by

$$\varphi(t) = (-y(t-1) \dots - y(t-n_a),\ u(t) \dots u(t-n_b))^T. \tag{2.1}$$

We will use the conventions

$$\varphi_0(t) = (-y_0(t-1) \dots - y_0(t-n_a),\ u_0(t) \dots u_0(t-n_b))^T \tag{2.2}$$

and

$$\tilde{\varphi}(t) = (-\tilde{y}(t-1) \dots - \tilde{y}(t-n_a),\ \tilde{u}(t) \dots \tilde{u}(t-n_b))^T \tag{2.3}$$

to denote the noise-free part and the noise-contribution of the regressor vector, respectively. It follows that

$$\varphi(t) = \varphi_0(t) + \tilde{\varphi}(t). \tag{2.4}$$

Further, we will introduce the extended versions of $\varphi(t)$ and θ as

$$\bar{\varphi}(t) = \begin{pmatrix} \varphi(t) \\ \underline{\varphi}(t) \end{pmatrix}, \quad \bar{\theta} = \begin{pmatrix} \theta \\ \underline{\theta} \end{pmatrix} \tag{2.5}$$

where $\underline{\varphi}(t)$ and $\underline{\theta}$ are to be defined later on. We make the following assumption.

A7. The noise part $\tilde{\underline{\varphi}}(t)$ of the extended regressor vector is uncorrelated with $\tilde{\varphi}(t)$ and $\tilde{y}(t)$:

$$R_{\tilde{\underline{\varphi}}\tilde{\varphi}} = 0, \quad r_{\tilde{\underline{\varphi}}\tilde{y}} = 0. \tag{2.6}$$

3 The BELS methods

The EIV system described by (1.1)-(1.3) can be expressed as a linear regressor model

$$y(t) = \varphi^T(t)\theta + \varepsilon(t), \tag{3.1}$$

where $\varepsilon(t) = \tilde{y}(t) - \tilde{\varphi}^T(t)\theta$. When the least squares (LS) method is applied to (3.1), the estimate $\hat{\theta}$ will be

$$\hat{\theta}_{LS} = R_{\varphi\varphi}^{-1} r_{\varphi y}, \tag{3.2}$$

which is biased due to the noise on both the input and the output sides. A bias-compensated least-squares (BCLS) scheme [5], is written as

$$\hat{\theta}_{BCLS} = \left(R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}}\right)^{-1} \left(r_{\varphi y} - r_{\tilde{\varphi}\tilde{y}}\right). \tag{3.3}$$

The basic idea of BCLS is to remove the noise-contribution parts from the covariance matrix $R_{\varphi\varphi}$ and the vector $r_{\varphi y}$ to get consistent estimates. The BELS algorithm is based on this principle.

In the compensated normal equation (3.3), the noise matrix $R_{\tilde{\varphi}\tilde{\varphi}}$ and the noise vector $r_{\tilde{\varphi}\tilde{y}}$ are constructed by the noise covariances of both the input and the output sides. For the cases with correlated output noise and white input noise, the detailed expressions of this noise matrix and vector are

$$\begin{aligned} R_{\tilde{\varphi}\tilde{\varphi}} &= \begin{pmatrix} R_{\tilde{y}} & 0 \\ 0 & R_{\tilde{u}} \end{pmatrix} \\ &= \begin{pmatrix} r_{\tilde{y}}(0) & \dots & r_{\tilde{y}}(n_a-1) & \\ \vdots & \ddots & \vdots & 0 \\ r_{\tilde{y}}(n_a-1) & \dots & r_{\tilde{y}}(0) & \\ & 0 & & r_{\tilde{u}}(0) I_{n_b+1} \end{pmatrix} \end{aligned} \tag{3.4}$$

$$r_{\tilde{\varphi}\tilde{y}} = \begin{pmatrix} r_{\tilde{y}}(1) & \dots & r_{\tilde{y}}(n_a) & 0_{1\times(n_b+1)} \end{pmatrix}^T. \tag{3.5}$$

Here, for a general noise process $x(t)$, we define its covariance function $r_x(\tau)$ as:

$$r_x(\tau) = E(x(\tau)x(t-\tau)), \quad \tau = 0, \pm 1, \pm 2, \dots .$$

The corresponding unknown noise parameter vector is defined as:

$$\rho = (\rho_u, \rho_y) = (r_{\tilde{u}}(0), r_{\tilde{y}}(0), r_{\tilde{y}}(1), \dots, r_{\tilde{y}}(n_a)). \tag{3.6}$$

In order to determine the noise covariances (and hence $R_{\tilde{\varphi}\tilde{\varphi}}$ and $r_{\tilde{\varphi}\tilde{y}}$), more relations are needed in addition to the equations (3.3). One such relation can be derived from the minimal value of the least squares criterion:

$$\begin{aligned} V_{LS} &= E[y(t) - \varphi^T(t)\hat{\theta}_{LS}]^2 \\ &= r_{\tilde{y}}(0) + \hat{\theta}_{LS}^T R_{\tilde{\varphi}\tilde{\varphi}}\theta - (\theta + \hat{\theta}_{LS})^T r_{\tilde{\varphi}\tilde{y}}, \end{aligned} \tag{3.7}$$

(see equation (43) in [8]). To get further relations for the noise covariances, an extended model structure is introduced in the BELS method. In [8], the model extension is done by appending an additional B parameter vector, leading to

$$\underline{\varphi}(t) = \begin{pmatrix} u(t - n_b - 1) \\ \vdots \\ u(t - n_b - p) \end{pmatrix}, \quad \underline{\theta} = \begin{pmatrix} b_{n_b+1} \\ \vdots \\ b_{n_b+p} \end{pmatrix}. \tag{3.8}$$

(What number of p to choose will be discussed later.) This choice does satisfy assumption **A7**.

Similar to (3.1), the linear EIV system (1.1)-(1.3) can also be represented by the extended model

$$y(t) = \bar{\varphi}^T(t)\bar{\theta} + \varepsilon(t). \tag{3.9}$$

Applying the ordinary LS estimate in this extended model leads to $\hat{\bar{\theta}}_{LS}$, which consists of two parts, $\hat{\theta}_{LS}$ and $\hat{\underline{\theta}}_{LS}$. By using the fact that the true value of $\underline{\theta}$ is a zero vector, the following relation is obtained as:

$$R_{\varphi\underline{\varphi}}^T R_{\varphi\varphi}^{-1}(R_{\tilde{\varphi}\tilde{\varphi}}\theta - r_{\tilde{\varphi}\tilde{y}}) = r_{\underline{\varphi}y} - R_{\varphi\underline{\varphi}}^T \hat{\theta}_{LS}. \tag{3.10}$$

See *Theorem 2* in [8] for details.

To sum up, the BELS algorithm consists of three equations:

- the compensated normal equations (3.3),
- the minimal value of the loss function (3.7),
- the additional equation (3.10) obtained by using an extended model.

In the BELS methods, we initially set $\hat{\theta} = \hat{\theta}_{LS}$, then solve equations (3.7) and (3.10) to get the estimates of the noise parameter vector ρ. Next, a new value of $\hat{\theta}$ is computed by using the compensated normal equation (3.3). This process is iterated until convergence.

4 Simplifying the third relation of BELS

As seen in the preceding section, the third relation (3.10) of the BELS algorithm is quite complex. In the following, we simplify this relation because it will not only make the computational process easy but also help us choosing the proper vector $\underline{\varphi}(t)$. The accuracy and convergency properties of the BELS algorithm were analyzed in [2] and [4]. The results therein show that the construction of the implemental vector $\underline{\varphi}(t)$, *i.e.* the type of the extended model, is important for achieving good estimates. In addition, the simplified algorithm will make the analysis of the properties of the BELS estimates much easier.

Reconsider the compensated normal equations (3.3). For the extended model, they will be rewritten as

$$
\begin{aligned}
&\left(\; R_{\bar{\varphi}\bar{\varphi}} - R_{\tilde{\bar{\varphi}}\tilde{\bar{\varphi}}} \; \right) \begin{pmatrix} \theta \\ 0 \end{pmatrix} = r_{\bar{\varphi}y} - r_{\tilde{\bar{\varphi}}\tilde{\bar{y}}} \\
\Leftrightarrow \; &\begin{pmatrix} R_{\varphi\varphi} - R_{\tilde{\varphi}\tilde{\varphi}} & R_{\varphi\underline{\varphi}} \\ R_{\underline{\varphi}\varphi} & R_{\underline{\varphi}\underline{\varphi}} - R_{\underline{\tilde{\varphi}}\underline{\tilde{\varphi}}} \end{pmatrix} \begin{pmatrix} \theta \\ 0 \end{pmatrix} = \begin{pmatrix} r_{\varphi y} - r_{\tilde{\varphi}\tilde{y}} \\ r_{\underline{\varphi}y} \end{pmatrix}.
\end{aligned}
$$

Looking at the lower block row gives

$$
R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y}. \tag{4.1}
$$

Proposition: The two relations (3.10) and (4.1) are equivalent.
proof: Substituting the relation (3.2) and using $R^T_{\varphi\underline{\varphi}} = R_{\underline{\varphi}\varphi}$ in (3.10) yields

$$
\begin{aligned}
&R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}(R_{\tilde{\varphi}\tilde{\varphi}}\theta - r_{\tilde{\varphi}\tilde{y}}) = r_{\underline{\varphi}y} - R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}r_{\varphi y} \\
\Leftrightarrow \; &R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}(R_{\tilde{\varphi}\tilde{\varphi}}\theta - r_{\tilde{\varphi}\tilde{y}} + r_{\varphi y}) = r_{\underline{\varphi}y} \\
\Leftrightarrow \; &R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}(R_{\tilde{\varphi}\tilde{\varphi}}\theta + r_{\varphi_0 y_0}) = r_{\underline{\varphi}y}.
\end{aligned} \tag{4.2}
$$

Then, by replacing $y_0(t)$ with $\varphi_0^T(t)\theta$, we get

$$
\begin{aligned}
&\text{Relation (3.10)} \\
\Leftrightarrow \; &R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}(R_{\tilde{\varphi}\tilde{\varphi}}\theta + R_{\varphi_0\varphi_0}\theta) = r_{\underline{\varphi}y} \\
\Leftrightarrow \; &R_{\underline{\varphi}\varphi}R^{-1}_{\varphi\varphi}R_{\varphi\varphi}\theta = r_{\underline{\varphi}y} \\
\Leftrightarrow \; &R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y} \\
\Leftrightarrow \; &\text{Relation (4.1)}.
\end{aligned} \tag{4.3}
$$

■

Remark 1: The simplified relation (4.1) can be further rewritten as:

$$
\begin{aligned}
&R_{\underline{\varphi}\varphi}\theta = r_{\underline{\varphi}y} \\
\Leftrightarrow \; &E\underline{\varphi}(t)(y(t) - \varphi^T(t)\theta) = 0 \\
\Leftrightarrow \; &E\underline{\varphi}(t)\varepsilon(t) = 0.
\end{aligned} \tag{4.4}
$$

Remark 2: The relation (4.1) is a form of IV equations. In the case of $p = n_a + n_b + 1$, it is exactly the basic IV approach for determining θ. Similarly to the instrument vector in the IV method, the vector $\underline{\varphi}(t)$ can be built in many ways as long as assumption **A7** holds.

Remark 3: The relation (4.1) is a system of *linear* equations, with only θ as unknowns. Since noise covariances are not present there, the computational load will be much lower than in (3.10).

Remark 4: To guarantee that the relation (4.1) will give p linearly independent equations, $R_{\underline{\varphi}\varphi}$ must have full rank.

Remark 5: The analysis carried out here covers also the case that $\tilde{u}(t)$ and $\tilde{y}(t)$ are both white noise. Then the additional vector $\underline{\varphi}(t)$ can be built by delayed inputs and/or delayed outputs.

Next, we analyze what value of p to be chosen. Consider that there are $2n_a + n_b + 3$ unknowns.

- $\theta{=}(a_1 \ldots a_{n_a}\ b_0 \ldots b_{n_b})^T \Rightarrow n_a{+}n_b{+}1$ unknowns
- $\rho = (r_{\tilde{u}}(0) r_{\tilde{y}}(0) \ldots r_{\tilde{y}}(n_a)) \Rightarrow n_a{+}2$ unknowns

And totally, there are $n_a + n_b + 2 + p$ equations.

- modified normal equations$\Rightarrow n_a{+}n_b{+}1$ equations
- minimal value of loss function $\Rightarrow$ 1 equation
- additional IV equations $\Rightarrow$ p equations

To have computability, the number of equations must be larger than or equal to the number of unknowns, namely,

$$\begin{aligned} & n_a + n_b + 2 + p \geq 2n_a + n_b + 3 \\ \Rightarrow \quad & p \geq n_a + 1. \end{aligned} \tag{4.5}$$

5 Conclusions

A simplified form of the BELS algorithm has been proposed for identifying the errors-in-variables systems, where the input is observed in the white noise and the output is observed in the colored noise. The result also fits the cases with white input noise and white output noise. The derived new relation of the BELS methods is a form of linear IV equations, which was proved to be equivalent to the previous relation proposed in [8], Theorem 2. By using the new relation, we can both reduce the computational load and simplify the analysis of the BELS methods.

Acknowledgment

This research was partially supported by The Swedish Research Council, contract 621-2005-4207, by a Research Grant from the Australian Research Council and by a Research Grant from the University of Western Sydney, Australia.

References

[1] S. Beghelli, R.P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.

[2] M. Hong, T. Söderström, and W. X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables identification. In *Proc. 14th IFAC Symposium on System Identification, SYSID 2006*, Newcastle, Australia, March 29-31 2006.

[3] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.

[4] T. Söderström, M. Hong, and W.X. Zheng. Convergence properties of bias-eliminating algorithms for errors-in-variables identification. *International Journal of Adaptive Control and Signal Processing*, 19:703–722, November 2005.

[5] P. Stoica and T. Söderström. Bias correction in least–squares identification. *International Journal of Control*, 35(3):449–457, 1982.

[6] J. K. Tugnait. Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control*, AC-37:476–485, 1992.

[7] W. X. Zheng. Transfer function estimation form noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, 12:365–380, 1998.

[8] W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47:1142–1147, 2002.

2006-040 Mei Hong, Torsten Söderström and Wei Xing Zheng: *A simplified form of the bias-eliminating least squares method for errors-in-variables identification*

2006-039 Andreas Hellander and Per Lötstedt: *Hybrid Method for the Chemical Master Equation*

2006-038 Markus Nordén, Henrik Löf, Jarmo Rantakokko, and Sverker Holmgren: *Geographical Locality and Dynamic Data Migration for OpenMP Implementations of Adaptive PDE Solvers*

2006-037 Elisabeth Larsson, Krister Åhlander, and Andreas Hall: *Multi-Dimensional Option Pricing using Radial Basis Functions and the Generalized Fourier Transform*

2006-036 Stefan Engblom: *A Discrete Spectral Method for the Chemical Master Equation*

2006-035 Owe Axelsson and Janos Karátson: *Mesh Independent Superlinear PCG Rates via Compact-Equivalent Operators*

2006-034 Agnes Rensfelt and Torsten Söderström: *Optimal Excitation for Nonparametric Identification of Viscoelastic Materials*

2006-033 Parosh Aziz Abdulla, Noomene Ben Henda, Richard Mayr, and Sven Sandberg: *Limiting Behavior of Markov Chains with Eager Attractors*

2006-032 Torsten Söderström: *Extending the Frisch Scheme for Errors-in-Variables Identification to Correlated Output Noise*

2006-031 Håkan Zeffer and Erik Hagersten: *A Case for Low-Complexity Multi-CMP Architectures*

2006-030 Lars-Henrik Eriksson: *The GTO Toolset and Method*

2006-029 Lars-Henrik Eriksson: *Use of Domain Theories in Applied Formal Methods*

2006-028 Ulrika Pettersson, Elisabeth Larsson, Gunnar Marcusson, and Jonas Persson: *Improved Radial Basis Function Methods for Multi-Dimensional Option Pricing*

2006-027 Jan Nordström: *Error Bounded Schemes for Time-Dependent Hyperbolic Problems*

2006-026 Jan Nordström, Ken Mattsson, and Charles Swanson: *Boundary Conditions for a Divergence Free Velocity-Pressure Formulation of the Incompressible Navier-Stokes Equations*

2006-025 Johan Wikström, Arvid Kauppi, Arne W. Andersson, and Bengt Sandblad: *Designing a Graphical User Interface for Train Traffic Control*

2006-024 Arvid Kauppi, Johan Wikström, Bengt Sandblad, and Arne W. Andersson: *Control Strategies for Managing Train Traffic, Difficulties Today and Solutions for the Future*

2006-023 Gergana Bencheva, Svetozar Margenov, and Jiří Starý: *MPI Implementation of a PCG Solver for Nonconforming FEM Problems: Overlapping of Communications and Computations*

2006-022 Niclas Sandgren and Petre Stoica: *On Moving Average Parameter Estimation*

2006-021 Gunilla Linde, Jonas Persson, and Lina von Sydow: *High-Order Adaptive Space-Discretizations for the Black–Scholes Equation*

UPPSALA
UNIVERSITET

August 2006
ISSN 1404-3203
http://www.it.uu.se/