

Separation techniques for errors-in-variables parameter estimation

Levente Hunyadi

Budapest University of Technology and Economics
Department of Automation and Applied Informatics
H-1111 Budapest, Goldmann György tér 3. IV. em.
e-mail: hunyadi@aut.bme.hu
web: www.aut.bme.hu/portal/hunyadi

Abstract: Errors-in-variables models are statistical models in which not only dependent (output) but also independent (input) variables are observed with error, i.e. they exhibit a symmetrical model structure in terms of errors. The application field for these models is diverse including speech and audio processing, signal processing, system identification, econometrics and time series analysis. Without explicit information on the ratio of input and output noise, however, it is in general not possible to identify a dynamic errors-in-variables system, i.e. to derive unique model parameter estimates. In this paper, we explore separation techniques that partition observations into characteristically different sets. Comparing model parameter estimates over the separated sets, it is possible to infer noise parameter estimates, resolving the unidentifiability issue.

1 Introduction

Errors-in-variables (EIV) models are statistical models in which not only dependent but also independent variables are observed with an error, i.e. they exhibit a symmetric model structure in terms of errors. As an example, consider the case in Figure 1 where a set of data has to be approximated with a linear model. A conventional least-squares solution to the problem (left) minimizes the error along the y -axis only, or in other words, the error is attributed to the y -component quantity only. In contrast, an errors-in-variables approach (right) is symmetric in the sense that data is treated to be observed with errors along both the x and y dimensions. Notice that the ratio of input and output noise variances (or graphically, the angle enclosed by the dotted lines and the vertical) is assumed to be known.

The application field for errors-in-variables models is diverse including computer vision, image reconstruction, speech and audio processing, signal processing, modal and spectral analysis, system identification, econometrics, time series analysis and astronomy [6]. They are motivated in several situations where the focus is on discovering, understanding or parameterizing the internal relationship between ob-

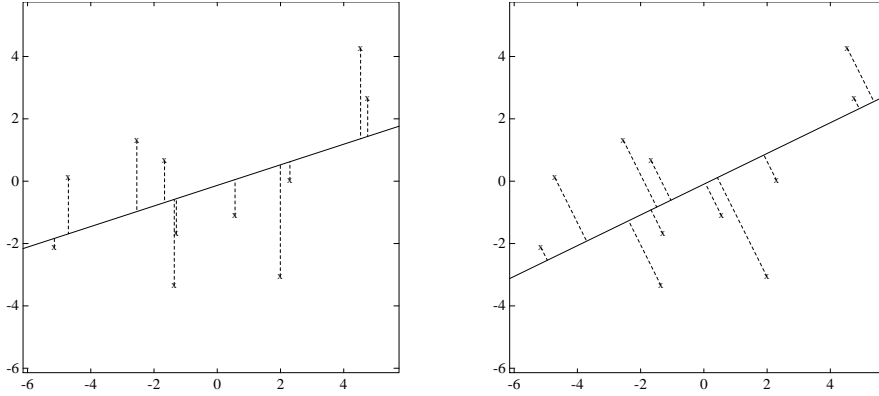


Figure 1: A conventional least-squares and an errors-in-variables approach to estimating a static linear parametric model.

served quantities. In these situations, an errors-in-variables approach usually yields more accurate models than conventional output-error only approaches. Errors-in-variables systems can be static (as the case in Figure 1) or dynamic.

For dynamic errors-in-variables systems, observations are not independent but there are couplings between observed quantities. Such a system is depicted in Figure 2, where a series of input data is fed to a process, yielding some output data. Supposing the system is linear, the process is described by a transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}$$

or by the equivalent autoregressive moving average (ARMA) linear difference equation

$$A(q^{-1})y_0(t) = B(q^{-1})u_0(t). \quad (1)$$

where q^{-1} denotes the backward shift operator such that $q^{-1}u(t) = u(t-1)$ and

$$\begin{aligned} A(q^{-1}) &= 1 + a_1q^{-1} + \dots + a_{ma}q^{-ma} \\ B(q^{-1}) &= b_1q^{-1} + \dots + b_{mb}q^{-mb}. \end{aligned}$$

Notice that only the noise-corrupted data sequences $u(t)$ and $y(t)$ are observable, the noise-free data sequences $u_0(t)$ and $y_0(t)$ are not. As far as the additive noise sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are concerned, a fairly reasonable assumption is to treat them as white Gaussian noise with variances σ_u^2 and σ_y^2 , respectively, which corresponds to noise due to measurement error. All samples are assumed to be equidistant in time. Although extensions to multi-input multi-output (MIMO) systems are possible, here we restrict ourselves to single-input single-output (SISO) systems. Without loss of generality, we may assume that $m = ma = mb$, replacing absent parameters with zeros, which allows us to use a symmetric model in terms

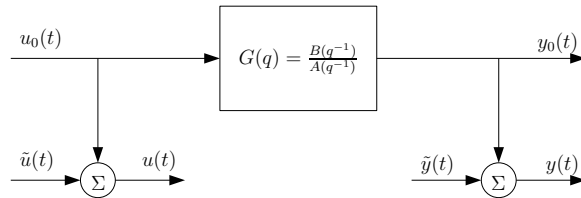


Figure 2: The basic setup for a discrete-time dynamic errors-in-variables system.

of model parameters. Given this system model, our goal is to derive *model* (or process) parameters a_i and b_i as well as *noise* parameters σ_u^2 and σ_y^2 using noise-contaminated observations $u(t)$ and $y(t)$, or in other words, *identify* a parametric system. The identification is performed using a parameter *estimator*.

There is extensive literature on the identification of parametric systems, see [8] for a comprehensive survey. Methods aiming at simultaneously deriving process and noise parameters include instrumental variables [9], bias-compensating least squares [13], the Frisch scheme [3, 5], structured total least squares [7], higher-order statistics [10], frequency-domain, prediction error and efficient maximum likelihood methods [12], which differ in the noise and experimental conditions they assume, the computational complexity they demand as well as the statistical accuracy they provide.

The fundamental problem in errors-in-variables identification is the deficiency in the knowledge of the noise variance ratio. In other words, the noise “direction” ρ that determines the extent to which observations are rather input or output noise laden is not available, that is, the angle at which the dotted lines in Figure 1 have to be constructed is unknown. The approach we chart in this paper is based on a preliminary separation step prior to estimating model and noise parameters. After an initial partitioning phase, separated sets are individually subject to parameter estimation, and the estimates are compared using a distance metric. Minimizing the distance metric over an assumed noise “direction” (or equivalently, a ratio of noise parameters), it is possible to arrive at a noise “direction” estimate. Once this information is at our disposal, the problem reverts to a classical identification problem, which can be tackled with conventional methods.

The key idea behind the aforementioned approach is that observations have a “hidden knowledge” of the true noise structure. The aim of the separation step is to partition the set of observations so that they are as far as possible from the perspective of the noise structure, i.e. they react differently to various assumptions of noise structure. Consequently, when the noise “direction” ρ is varied and parameter estimates are derived for each value of ρ , they are likely to differ substantially when an incorrect “direction” has been assumed. On the other hand, if the assumption for ρ matches the true value, the two sets of observations are likely to behave similarly when subject to parameter estimation.

2 Setup and notations

Given the system description (1), we may introduce the model parameter vector θ as well as its autoregressive and moving average components, θ^a and θ^b , respectively:

$$\begin{aligned}\theta &= [a_1 \quad \dots \quad a_{ma} \quad -b_1 \quad \dots \quad -b_{mb}]^\top \\ \theta^a &= [a_1 \quad \dots \quad a_{ma}]^\top \\ \theta^b &= [b_1 \quad \dots \quad b_{mb}]^\top\end{aligned}$$

whose estimates are denoted by $\hat{\theta}$ and whose true values by θ_0 . In general, the notation $\hat{\cdot}$ and \cdot_0 will also be applied to other parameters to indicate the estimated and the true value, respectively.

Similarly, the regressor vector $\varphi(t)$ may be introduced as

$$\begin{aligned}\varphi(t) &= [\varphi_y^\top(t) \quad \varphi_u^\top(t)]^\top \\ \varphi_y(t) &= [y(t-1) \quad \dots \quad y(t-ma)]^\top \\ \varphi_u(t) &= [u(t-1) \quad \dots \quad u(t-mb)]^\top\end{aligned}$$

so that the system description in (1) can be reformulated in the compact linear regression form

$$y(t) = \varphi^\top(t)\theta + \varepsilon(t) \quad (2)$$

where ε is a stochastic disturbance term $\varepsilon(t) = \tilde{y}(t) - \tilde{\varphi}^\top(t)\theta_0$ in which $\tilde{\varphi}$ is the noise contribution of the regressor vector.

In order to emphasize the symmetric nature of the errors-in-variables approach, it is preferable to exploit the symmetry of EIV models and use an implicit formula rather than the explicit formula in (2). For this end, supplement the model parameters in θ with additional elements such that

$$g = [a_0 \quad a_1 \quad \dots \quad a_m \quad -b_0 \quad -b_1 \quad \dots \quad -b_m]^\top$$

and write

$$x^\top(t)g = 0 \quad (3)$$

where

$$\begin{aligned}x(t) &= [x_y^\top(t) \quad x_u^\top(t)]^\top \\ x_y(t) &= [y(t) \quad \dots \quad y(t-m)]^\top \\ x_u(t) &= [u(t) \quad \dots \quad u(t-m)]^\top\end{aligned}$$

for $t = 0, \dots, N - m$ where the implicit assumptions $a_0 = 1, b_0 = 0$ have been made to make (3) conform to (2).

In many cases, it is more practical to use matrix notation by collecting multiple observations into a large vector or matrix. Notations such as u or y refer to these

N -row vectors, while Φ and X collect $N - m + 1$ and $N - m$ observations of $\varphi(t)$ and $x(t)$, respectively.

$$\begin{aligned} u &= [u_1 \ u_2 \ \dots \ u_N]^\top \\ y &= [y_1 \ y_2 \ \dots \ y_N]^\top \\ \Phi &= \begin{bmatrix} y_m & \dots & y_1 & u_m & \dots & u_1 \\ y_{m+1} & \dots & y_2 & u_{m+1} & \dots & u_2 \\ \vdots & & \vdots & \vdots & & \vdots \\ y_N & \dots & y_{N-m+1} & u_N & \dots & u_{N-m+1} \end{bmatrix} \\ X &= \begin{bmatrix} y_{m+1} & \dots & y_1 & u_{m+1} & \dots & u_1 \\ y_{m+2} & \dots & y_2 & u_{m+2} & \dots & u_2 \\ \vdots & & \vdots & \vdots & & \vdots \\ y_N & \dots & y_{N-m} & u_N & \dots & u_{N-m} \end{bmatrix} \end{aligned}$$

In addition, it is possible to define matrices with a greater number of past observations. Accordingly, matrices such as Φ_q or X_q may be introduced that take $q - 1$ or q instead of m or $m + 1$ columns per input or output.

With matrix notation, (2) can be concisely written as an overdetermined system of equations (dropping the first few rows of y such that the dimensions match)

$$y = \Phi\theta + \varepsilon.$$

As far as noise assumptions are concerned, the covariance matrix of white input–output measurement noise is parameterizable with two scalars: μ corresponding to noise magnitude, and ρ to noise “direction”, such that

$$C = \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} = \mu C_\rho = \mu \begin{bmatrix} \sin^2 \rho & 0 \\ 0 & \cos^2 \rho \end{bmatrix}. \quad (4)$$

Likewise, observations can be characterized with their sample covariance matrices. Define the sample covariance matrices and vectors R_φ and $r_{\varphi y}$, as well as their estimates \hat{R}_φ and $\hat{r}_{\varphi y}$, in a way that

$$\begin{aligned} R_\varphi &= \mathbb{E}(\varphi(t)\varphi^\top(t)) \\ r_{\varphi y} &= \mathbb{E}(\varphi(t)y(t)) \\ \hat{R}_\varphi &= \frac{1}{N} \sum_{t=1}^N \varphi(t)\varphi^\top(t) = \frac{1}{N} \Phi^\top \Phi \\ \hat{r}_{\varphi y} &= \frac{1}{N} \sum_{t=1}^N \varphi(t)y(t) = \frac{1}{N} \Phi^\top y \end{aligned}$$

where \hat{R}_φ and $\hat{r}_{\varphi y}$ are estimates for R_φ and $r_{\varphi y}$ from N samples. A similar covariance matrix R may be introduced for the observation vector $x(t)$, given in the

implicit form (3), which incorporates the covariance matrix for both $\varphi(t)$ and $y(t)$. Similarly, the respective correlation matrices ϕ may also be defined such that for each entry (i, j) of ϕ

$$\phi_{ij} = \frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}}.$$

3 Estimation with preliminary separation

The idea of applying a preliminary separation step to partition observation has been used in [7] for static systems with some limited extensions to dynamic systems. In the static case, a k-means clustering algorithm is employed to split observations into disjoint groups. An input-to-output noise variance ratio is assumed, and parameter estimates are derived using the total least squares approach [1] for each ratio. Next, the estimates are compared using some distance metrics. Whenever these metrics measure minimum, they conclude that the “true” noise ratio has been found.

In contrast, for the dynamic case they apply the structured version of total least squares, which does exploit the sequential arrangement of observations and hence does not permit a clustering algorithm. Consequently, the authors have assumed that data is split into two *contiguous* blocks of observations, which, they admit, is very restrictive in practice. In fact, such identification approach can be discussed in a repeated experiments framework, in which model and noise properties remain the same throughout the experiments.

The approach presented here does not make such assumptions and can operate on *non-contiguous* blocks of observations. The estimation procedure takes the following steps:

1. A *noise structure* is assumed. As estimator methods often automatically compute noise magnitude given a noise covariance structure, it is sufficient to parameterize a covariance matrix C in (4) corresponding to white noise with a single scalar ρ that represents noise “direction”.
2. Using the noise-polluted observations $y(t)$ and $u(t)$, a possibly extended observation matrix X_q is constructed, where q is a parameter of the user’s choice, usually with $q \gg m$.
3. A *transformation* is applied on X_q . The goal of the transformation is to emphasize the characteristics based on which the separation is performed.
4. Rows of X_q , each of which represents an observation at time t , are grouped into two sets by means of a *separation algorithm*.
5. The *set estimator* derives parameter estimates for each of the sets independently by a loss function given the chosen noise model.
6. Parameter estimates for the two sets are compared using some *distance metrics*.

As the value of ρ is within the range $[0; \frac{\pi}{2}]$ (0 corresponding to input noise, $\frac{\pi}{2}$ to output noise only), minimizing the distance metrics yields the “true” value for $\hat{\rho}$. Once an estimate for ρ is at our disposal, we may apply an efficient maximum likelihood estimator [12] to compute “true” model parameters estimates $\hat{\theta}$ as well as the noise magnitude $\hat{\mu}$, and hence $\hat{\sigma}_y$ and $\hat{\sigma}_u$.

Notice the underlying assumption that separation should produce sets with sufficiently different characteristics. Otherwise, estimates may be close to each other even if the noise structure is not appropriate, yielding a false value for noise variances and, in turn, model parameter values. It is therefore of paramount importance to select a proper separation algorithm, which we discuss in a later section.

3.1 The generalized Koopmans–Levin estimator

Our approach uses the generalized Koopmans–Levin algorithm [11] to compute model parameter estimates given a noise structure. This method minimizes the loss function

$$J = \frac{1}{2} \text{tr} \left((G_q^\top C_q G_q)^{-1} G_q^\top D_q G_q \right). \quad (5)$$

where

$$G_q = \begin{bmatrix} G_q^a \\ -G_q^b \end{bmatrix}$$

in which G_q^a and G_q^b are banded Toeplitz matrices of parameters a_i and b_i such that $(X_0)_q G_q = 0$:

$$G_q^a = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ a_1 & 1 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_m & a_{m-1} & a_{m-2} & \dots & 0 & 0 \\ 0 & a_m & a_{m-1} & \dots & 0 & 0 \\ 0 & 0 & a_m & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_m & a_{m-1} \\ 0 & 0 & 0 & \dots & 0 & a_m \end{bmatrix}_{q, q-m}$$

and G_q^b can be constructed in a similar manner. Notice that the construction is analogous to (3). The covariance matrix C_q is constructed as $C_q = (\mu C_\rho) \otimes I_q$ and $D_q = X_q^\top X_q$. The parameter q in (5) is of the user’s choice such that $q \gg m + 1$, with higher values (to a limit) yielding more accurate results at the expense of computational cost. The special choices $q = m + 1$ yields the original Koopmans–Levin algorithm, while $q = N$ produces the maximum likelihood estimator, hence q can be seen as a scaling parameter. In simulation examples, we choose q such that $15 \leq q \leq 60$.

Notice that C_q in (5) is a known parameter up to multiplication by a scalar because we have already committed ourselves to a given noise structure before invoking the estimation algorithm. In contrast, G_q wraps the unknown model parameters, so that only iterative procedures can be sought in order to minimize J . We employ the Fletcher version [4] of the Levenberg–Maquardt search algorithm to minimize J in (5).

While we have selected the generalized Koopmans–Levin estimator, it is equally possible to use other types of estimators that need a noise structure model and can tackle a discontinuous separation of observations.

3.2 Observation separation techniques

The goal of data separation is to devise an unsupervised analysis to partition observations into disjoint sets such that points belonging to the same set are similar, while those belonging to different sets are dissimilar. This can be formalized by introducing a set indicator so that

$$\begin{aligned} S_1 &= \{i \mid f(x_q(i)) \geq 0\} \\ S_2 &= \{i \mid f(x_q(i)) < 0\} \end{aligned}$$

where $x_q(i)$ denotes the i th row of X_q and $f, : \mathbb{R}^q \rightarrow \mathbb{R}$, $i = 1 \dots N$ such that $S_1 \cap S_2 = \emptyset$.

The most natural way to assess the performance of the function f is to compare the covariance matrices R_1 and R_2 the separated observations it brings forth would produce. The aim is to produce characteristically different elements in the lower (or equivalently, upper) triangle of R_1 and R_2 calculated by taking the observations that belong to each of the two respective sets. This can be numerically measured using matrix divergences. Let ϕ_1 and ϕ_2 denote the correlation matrices of the respective sets. The Itakura–Saito matrix divergence is computed as

$$d_{\phi_1} = \text{trace}(\phi_1 \phi_2^{-1}) - \log(\det(\phi_1 \phi_2^{-1})) - n$$

where n is the order of the correlation matrix ϕ , whereas the von Neumann matrix divergence is calculated as

$$d_{\phi_1} = \text{trace}(\phi_1 \log \phi_1 - \phi_1 \log \phi_2 - \phi_1 + \phi_2)$$

where $\log X$ is the principal matrix logarithm of X . The index emphasizes that these divergences are not symmetric, in order to make them independent of the substitution order of ϕ_1 and ϕ_2 , we may use $d = \max(d_{\phi_1}, d_{\phi_2})$. However, it is clear that maximizing d is a cumbersome endeavor even with a greedy algorithm. Consequently, computationally simpler alternatives have to be considered. One idea is to transform the original observation matrix X_q using a transformation T and do the separation over TX_q .

Principal component analysis (PCA) is a widely used statistical method for dimension reduction. The basis for dimension reduction is that PCA picks up the

dimensions with the largest variances. The idea of PCA-based separation is to compute the singular value decomposition (SVD), which is the basis for PCA, and inspect one or more of the principal singular vectors. More specifically, decompose the data matrix \bar{D} such that

$$\bar{D} = \bar{U}\Sigma\bar{V}^\top$$

and denote the columns of \bar{U} as \bar{u}_i so that the first principal vector is \bar{u}_1 . A set indicator may then be introduced so that

$$\begin{aligned} S_1 &= \{i \mid f(\bar{u}_{p_i}(i), \dots, \bar{u}_{p_f}(i)) \geq 0\} \\ S_2 &= \{i \mid f(\bar{u}_{p_i}(i), \dots, \bar{u}_{p_f}(i)) < 0\} \end{aligned}$$

where $f : \mathbb{R}^{p_f - p_i} \rightarrow \mathbb{R}$, $i = 1 \dots N$ and $p_f - p_i$ determines how many principal components to take into consideration. Choices to f include:

- $u_1(i) \geq 0$ which is essentially equivalent to performing a k -means clustering on the data with $k = 2$ (see [2]).
- $\prod_{k=p_i}^{p_f} \bar{u}_k(i) \geq 0$. If corresponding elements in the covariance matrices have opposite signs, it is likely that the estimation algorithm produces similar estimates for the two sets only in case of correct noise assumption. A natural combination is to choose $p_i = 1$ and $p_f = 2$.
- $|\bar{u}_1|(i) > m_1$ where m_1 is the median of the values in the first principal vector \bar{u}_1 .
- $\|\bar{u}_{p_i \dots p_f}\| > m$, which is a generalization of the above, where $\bar{u}_{p_i \dots p_f}$ stands for the indexed principal components and m is the median value of the norm. For a 2-dimensional case with $p_i = 1$ and $p_f = 2$, this corresponds to a circle in the \bar{u}_1 vs. \bar{u}_2 plane where observations are grouped whether they fall inside or outside the circle.

What remains to discuss is the exact matrix to use in place of the data matrix \bar{D} that is subject to decomposition. The following are viable alternatives:

- the extended observation matrix X_q ; or
- the components of X_q that correspond to input observations, which we denote by U_q .

Notice that the observation matrix X_q consists of both input and output observations, each of which is contaminated with noise with a different variance σ_u^2 and σ_y^2 , respectively. Consequently, it is better to replace the Euclidean distance with the Mahalanobis distance that takes the different scalings into account by incorporating the noise matrix $\bar{C} = C_q$ in (5) into separation mechanisms. In accordance,

the generalized version of singular value decomposition (GSVD) has to be employed instead of SVD such that

$$\begin{aligned}\bar{D} &= \bar{U}\Sigma_1\bar{X}^\top \\ \bar{C} &= \bar{V}\Sigma_2\bar{X}^\top \\ I &= \Sigma_1^\top\Sigma_1 + \Sigma_2^\top\Sigma_2\end{aligned}$$

where \bar{U} and \bar{V} are unitary matrices and I is the unit matrix.

Obviously, principal components analysis is not the only alternative in transforming the observation matrix X_g . In particular, applying a discrete (possibly short-time) Fourier-transform to the observation vector y or u and performing the separation in (partially) frequency rather than (purely) time domain is a viable alternative. In this scenario, the separated sets will not be disjoint in time domain even though each set will contain only selected frequency components. A simple (though not very effective) frequency domain separation is to split the spectrum into low-frequency and high-frequency components.

3.3 Comparing parameter estimates

There are various ways parameter estimates over the separated sets can be compared. The most straightforward is to use the relative distance

$$d = \frac{\|\hat{\theta}_1 - \hat{\theta}_2\|}{\|\hat{\theta}_1\| \|\hat{\theta}_2\|}$$

where $\hat{\theta}_k$ represents the estimated parameter vector on set k . It is, however, often more practical to compare only autoregressive (AR) components $\hat{\theta}_k^a$ of the model, i.e. parameters a_i , which often produces more accurate results, especially for sequences with low moving average excitation. Accordingly,

$$d = \frac{\|\hat{\theta}_1^a - \hat{\theta}_2^a\|}{\|\hat{\theta}_1^a\| \|\hat{\theta}_2^a\|}. \quad (6)$$

As a third option, the angle enclosed by the estimated parameter vectors may be compared, such that

$$d = \angle(\theta_1, \theta_2).$$

where \angle denotes the angle enclosed by the model parameter vectors.

These metrics do not take noise magnitude into account. The combined distance metrics

$$d = ((\mu_1 - \mu_2)^2 + c \sin^2(\angle(\theta_1, \theta_2)))$$

proposed in [7] can be leveraged for a possibly more accurate noise direction estimate where c is a scaling constant, often chosen as $c = 1$.

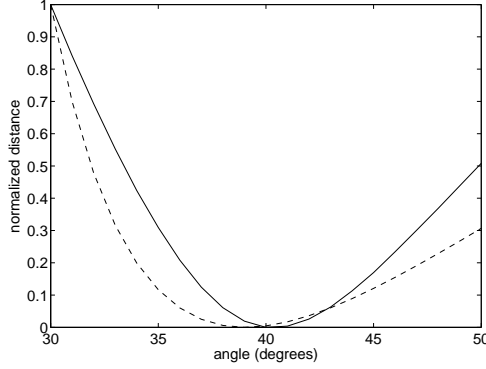


Figure 3: Minimizing the relative parameter distance metric over a noise “direction” measure. The separation was performed using a greedy maximization of the Itakura–Saito matrix divergence (continuous) and principal component analysis (dashed).

4 Simulation results

Consider the discrete linear model described by the relationship

$$y_0(t) = \frac{B(q^{-1})}{A(q^{-1})} u_0(t) = 0.05 \frac{2q^{-1} + q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}} u_0(t) \quad (7)$$

and let $N = 2048$, $\rho = 40^\circ$, $\mu = 0.1$. The input sequence sampled from a composite sinusoidal signal and the amplifier coefficient 0.05 have been chosen to produce a signal-to-noise (SNR) ratio of approximately 10dB where SNR is measured by

$$SNR[dB] = 10 \log_{10} \frac{\frac{1}{N} \sum_{i=1}^N u_0^2}{\frac{1}{N} \sum_{i=1}^N (u - u_0)^2}$$

As far as the parameters of the identification algorithm are concerned, let $q = 16$ in the GKL algorithm, use the Itakura–Saito matrix divergence and principal component analysis (the latter with set indicator $\bar{u}_1(i)\bar{u}_2(i) \geq 0$) in separating the entire observation matrix X_q and measure the distance between the model parameter estimates using the relative distance (6) based on autoregressive components only. The Itakura–Saito divergence was maximized using a greedy algorithm. First, both sets were initialized with the entire set of observations after which observations were alternately omitted from either set such that the divergence would increase. Figure 3 shows that the distance between model parameter estimates is indeed minimized at the true value of the noise parameter as a noise “direction” measure is varied in the range $[0; \frac{\pi}{2}]$.

5 Conclusions

We have investigated an approach to identifying linear dynamic errors-in-variables systems with a preliminary separation step. We have seen that the aim of the separation step is to produce two distinct sets which are distant from each other in a certain sense. In other words, when parameter estimates are derived for each of the two sets, they are likely to be close to one another only if an initial noise assumption was correct. In fact, assuming an incorrect noise covariance structure leads to easily identifiable groups of observations, whereas a correct assumption makes no such distinction of observations possible. As a result, traversing a noise space, the “true” noise model can be discovered by minimizing the distance between parameter estimates over the two sets. Future work is focused on identifying efficient separation mechanisms both in time and frequency domain, or alternatively, based on other transformed versions of the original observation matrix.

References

- [1] Peter de Groen. An introduction to total least squares. *Nieuw Archief voor Wiskunde*, 4(14):237–253, 1996.
- [2] Chris Ding and Xiaofeng He. K-means clustering via principal component analysis. In *Proceedings of the 21st International Conference on Machine Learning*, page 29, New York, NY, USA, 2004. ACM.
- [3] Roberto Diversi, Roberto Guidorzi, and Umberto Soverini. A new criterion in EIV identification and filtering applications. In *Proc. of 13th IFAC Symposium on System Identification*, pages 1993–1998, 2003.
- [4] R. Fletcher. A modified Marquardt subroutine for nonlinear least squares. Report No. R-6799, Atomic Energy Research Estab., Harwell, Berkshire, 1971.
- [5] Mei Hong, Torsten Söderström, Umberto Soverini, and Roberto Diversi. Comparison of three Frisch methods for errors-in-variables identification. In *Proc. of 17th IFAC World Congress*, pages 420–425, Seoul, Korea, July 2008.
- [6] Sabine Van Huffel and Philippe Lemmerling, editors. *TLS and EIV modeling*. Kluwer, 2002.
- [7] Ivan Markovsky, Alexander Kukush, and Sabine Van Huffel. On errors-in-variables with unknown noise variance ratio. In *Proc. of 14th IFAC Symposium on System Identification*, pages 172–177, March 2006.
- [8] Torsten Söderström. Errors-in-variables methods in system identification. *Automatica*, 43:939–958, 2007.
- [9] Stéphane Thil, Marion Gilson, and Hugues Garnier. On instrumental variable-based methods for errors-in-variables model identification. In *Proc. of 17th IFAC World Congress*, pages 420–425, Seoul, Korea, July 2008.
- [10] Stéphane Thil, Mei Hong, Torsten Söderström, Marion Gilson, and Hugues Garnier. Statistical analysis of a third-order cumulants based algorithm for discrete-time errors-in-variables identification. In *Proc. of 17th IFAC World Congress*, pages 420–425, Seoul, Korea, July 2008.
- [11] István Vajk. Identification methods in a unified framework. *Automatica*, 41(8):1385–1393, 2005.
- [12] István Vajk and Jenő Hetthéssy. Efficient estimation of errors-in-variables models. In *17th IFAC World Congress*, Seoul, Korea, July 2008.
- [13] Wei Xing Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, July 2002.