INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2006; **20**:337–351
Published online 21 April 2006 in Wiley InterScience (www.interscience.wiley.com) DOI:10.1002/acs.905

Least-correlation estimates for errors-in-variables models

Byung-Eul Jun[1,*,†] and Dennis S. Bernstein[2,‡]

[1] *Agency for Defense Development, P.O. Box. 35-3 Youseong, Daejeon 305-600, Korea*
[2] *Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, U.S.A.*

SUMMARY

This paper introduces an estimator for errors-in-variables models in which all measurements are corrupted by noise. The necessary and sufficient condition minimizing a criterion, defined by squaring the empirical correlation of residuals, yields a new identification procedure that we call least-correlation estimator. The method of least correlation is a generalization of least-squares since the least-correlation specializes to least-squares when the correlation lag is zero. The least-correlation estimator has the ability to estimate true parameters consistently from noisy input–output measurements as the number of samples increases. Monte Carlo simulations also support the consistency numerically. We discuss the geometric property of the least-correlation estimate and, moreover, show that the estimate is not an orthogonal projection but an oblique projection. Finally, recursive realizations of the procedure in continuous-time as well as in discrete-time are mentioned with a numerical demonstration. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: parameter estimation; identification; least correlation; least squares; errors-in-variables model; input–output noise

1. INTRODUCTION

Although method of least-squares developed in the 18th century is still the most popular approach for determining the best fit to a given structure [1], this technique exhibits high sensitivity to errors in regressors [2, 3]. A generalized approach to modelling noise is to view all variables as contaminated by noise, called errors-in-variables (EIV) models [4–10]. These models have broad application in time series modelling, image processing, signal processing, neural networks and system identification in the fields of engineering, econometrics, and statistics. A recent trend in systems is the use of powerful computers with low-grade instruments, which implies that the needs for EIV models are increasing.

We introduce a criterion defined by squaring the empirical autocorrelation of residuals and derive an estimate minimizing the criterion with sufficiency and necessity. Optimality in the sense of least correlation has several attractive features. The approach works not only on static systems but also on dynamic systems since it does not depend on the specific structure of the

*Correspondence to: B.-E. Jun, Agency for Defense Development, P.O. Box. 35-3 Youseong, Daejeon 305-600, Korea.
† E-mail: bejun@add.re.kr
‡ E-mail: dsbaero@umich.edu

Received 10 February 2005
Revised 7 February 2006
Accepted 28 February 2006

Copyright © 2006 John Wiley & Sons, Ltd.

regressors. Only simple matrix algebra is needed to derive the estimate. There is a direct relationship between the criterion and the stochastic correlation. It works on EIV models with coloured noises as well as with white noises. The literature reveals that the 'error whitening Wiener filter' [11, 12] can be considered as a stochastic counterpart of the least-correlation estimate introduced in this work. Under the setting in References [11, 12], the estimation residuals are whitened by the 'error whitening Wiener filter'. In the generalized formulation of the present paper, however, the whitening filter is not able to whiten the residuals any more. In this sense, the term 'error whitening' cannot be used for this work.

In Section 2, we set the system model, define the problem and introduce assumptions. Section 3 begins with the definition of the least-correlation criterion and proceeds to the derivation of the least-correlation estimate. The estimate satisfies sufficiency and necessity in order to minimize the criterion. Consistency of the estimate is also mentioned. We discuss in Section 4 the geometrical aspect of the new estimate. Its geometrical property is partly similar to and partly different from that of the least-squares. An example shows that the least-correlation estimate is not an orthogonal projection but a kind of oblique projection. Section 5 shows a numerical example. Monte Carlo simulations support the consistency numerically. Section 6 states the recursive least-correlation (RLC) realization of the least-correlation estimator and shows a numerical demonstration of the RLC algorithm. Concluding remarks are given in Section 7.

2. PROBLEM FORMULATION

Consider the linear regression model

$$z(t) = \phi^{\mathrm{T}}(t)\theta + \eta_1(t) \tag{1}$$

where $z(t) \in \mathbb{R}$ is the system response at the tth sample time, $\phi(t) \in \mathbb{R}^n$ is the regression vector, $\theta \in \mathbb{R}^n$ is the parameter vector and $\eta_1(t) \in \mathbb{R}$ denotes possible residuals in modelling. The components of $\phi(t)$ depend on the type of system models. For example, $\phi(t)$ is composed of current inputs in linear static systems, delayed inputs in finite impulse response (FIR) systems, delayed outputs in all-pole systems or in autoregression (AR) models, and both delayed inputs and delayed outputs in autoregression with exogenous variables (ARX) models.

Let $y(t)$ and $\psi(t)$ denote noisy measurements of $z(t)$ and $\phi(t)$ as

$$y(t) = z(t) + \eta_2(t) \tag{2}$$

$$\psi(t) = \phi(t) + \zeta(t) \tag{3}$$

where $\eta_2(t) \in \mathbb{R}$ and $\zeta(t) \in \mathbb{R}^n$ are additive noises. Taking into account the noisy measurements of regressors as well as outputs constitutes an EIV problem [5, 8–10, 13]. Applying (2), (3) to (1) yields

$$y(t) = \psi^{\mathrm{T}}(t)\theta + e(t) \tag{4}$$

$$e(t) = \eta(t) - \zeta^{\mathrm{T}}(t)\theta \tag{5}$$

where $\eta(t) \triangleq \eta_1(t) + \eta_2(t)$ denotes the total error on output measurements. Let us state the identification problem for the EIV model.

Copyright © 2006 John Wiley & Sons, Ltd.

Problem 1
Given the system model (1) and the measurement model (2), (3), estimate the system parameter θ based on the available measurements $\psi(t)$ and $y(t)$.

For a common framework for deterministic and stochastic signals [13, p. 34], we employ the definition of quasi-stationary signals and the notation

$$\bar{E}[f(t)] \triangleq \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} E[f(t)] \tag{6}$$

which works on the deterministic components as well as the stochastic parts of the quasi-stationary signal $f(t)$, where E denotes mathematical expectation. We implicitly assume that the limit in (6) exists when $\bar{E}$ is used.

We introduce the following assumptions.

A1. The system is represented as a linear regression model, and the number of parameters to be estimated is known *a priori*. If the system is dynamic, it is uniformly stable.
A2. Measurements $\psi(t)$ and $y(t)$ are jointly quasi-stationary.
A3. The noises $\eta(t)$ and $\zeta(t)$ are zero mean and at most finitely cross-correlated with $\psi(t)$, that is, there exists $\tau > 0$ such that

$$\bar{E}[\phi(t)\zeta^{\mathrm{T}}(t-k)] = 0, \quad \bar{E}[\zeta(t)\zeta^{\mathrm{T}}(t-k)] = 0 \tag{7}$$

$$\bar{E}[\phi(t)\eta(t-k)] = 0, \quad \bar{E}[\zeta(t)\eta(t-k)] = 0 \tag{8}$$

for all $|k| \geqslant \tau$.
A4. For τ in A3, $\psi(t)$ satisfies

$$\operatorname{rank}[\bar{R}_{\psi\psi}(t, t-\tau, N) + \bar{R}_{\psi\psi}(t-\tau, t, N)] = n \tag{9}$$

where N is the number of samples and the empirical correlation matrix $\bar{R}_{\psi\psi}(t_1, t_2, N)$ with $t_1 = t$, $t_2 = t - \tau$ or $t_1 = t - \tau$, $t_2 = t$ is defined by

$$\bar{R}_{\psi\psi}(t_1, t_2, N) \triangleq \frac{1}{N - |t_1 - t_2|} \sum_{t=1+|t_1-t_2|}^{N} \psi(t_1)\psi^{\mathrm{T}}(t_2) \tag{10}$$

3. LEAST-CORRELATION ESTIMATES

Consider an arbitrary estimate $\bar{\theta}$ with the residual

$$\varepsilon(t, \bar{\theta}) = y(t) - \psi^{\mathrm{T}}(t)\bar{\theta} \tag{11}$$

which models the mismatch between observations and estimates. If the residual from the least-squares estimate is not white, then at least one of the following statements is true.

(1) Either the modelling is not complete or the estimation procedure has an error at least.
(2) Either $\eta(t)$ or $\zeta(t)$ is coloured.
(3) There is a non-zero correlation between $\psi(t)$ and either $\eta(t)$ or $\zeta(t)$.

Copyright © 2006 John Wiley & Sons, Ltd.

Applying (4), (5) to (11) gives conditions (1) and (2). Condition (3) is obtained by applying least-squares to the EIV problems (4), (5) [2, 14].

Given $\bar{\theta} = \theta$ and $\eta(t) = 0$, the correlation $\bar{E}[\varepsilon(t,\bar{\theta})\varepsilon(t',\bar{\theta})]$ is zero for all $|t - t'| \geqslant \tau$ due to A3, but the mean square error $\bar{E}[\varepsilon^2(t,\bar{\theta})]$ is never zero. Based on this insight, we introduce a criterion

$$J^2(\bar{\theta},\tau,N) = \left(\frac{1}{N_\tau}\sum_{t=1+\tau}^{N} \varepsilon(t,\bar{\theta})\varepsilon(t-\tau,\bar{\theta})\right)^2 \tag{12}$$

where $N_{g(\tau)} \triangleq N - g(\tau)$, τ is an integer defining the correlation lag and N denotes the number of samples. It is noted that $J(\bar{\theta},0,N)$ is the criterion of least squares. Function (12) is rewritten as

$$J^2(\bar{\theta},\tau,N) = \left\{\frac{1}{2N_\tau}(Y_{0/\tau} - \Psi_{0/\tau}\bar{\theta})^{\mathrm{T}}(Y_{\tau/0} - \Psi_{\tau/0}\bar{\theta})\right\}^2 \tag{13}$$

with the stacks of measurements $\Psi_{0/\tau}$, $\Psi_{\tau/0} \in \mathbb{R}^{2N_\tau \times n}$ and $Y_{0/\tau}$, $Y_{\tau/0} \in \mathbb{R}^{2N_\tau}$ defined by

$$\Psi_{0/\tau} \triangleq \begin{bmatrix} \Psi_0 \\ \Psi_\tau \end{bmatrix}, \quad \Psi_{\tau/0} \triangleq \begin{bmatrix} \Psi_\tau \\ \Psi_0 \end{bmatrix}, \quad Y_{\tau/0} \triangleq \begin{bmatrix} Y_\tau \\ Y_0 \end{bmatrix}, \quad Y_{0/\tau} \triangleq \begin{bmatrix} Y_0 \\ Y_\tau \end{bmatrix} \tag{14}$$

where Y_0, $Y_\tau \in \mathbb{R}^{N_\tau}$ and Ψ_0, $\Psi_\tau \in \mathbb{R}^{N_\tau \times n}$ are defined by

$$Y_0(\tau,N) \triangleq \begin{bmatrix} y(N) \\ y(N-1) \\ \vdots \\ y(1+\tau) \end{bmatrix}, \quad Y_\tau(\tau,N) \triangleq \begin{bmatrix} y(N_\tau) \\ y(N_{\tau+1}) \\ \vdots \\ y(1) \end{bmatrix} \tag{15}$$

$$\Psi_0(\tau,N) \triangleq \begin{bmatrix} \psi^{\mathrm{T}}(N) \\ \psi^{\mathrm{T}}(N-1) \\ \vdots \\ \psi^{\mathrm{T}}(1+\tau) \end{bmatrix}, \quad \Psi_\tau(\tau,N) \triangleq \begin{bmatrix} \psi^{\mathrm{T}}(N_\tau) \\ \psi^{\mathrm{T}}(N_{\tau+1}) \\ \vdots \\ \psi^{\mathrm{T}}(1) \end{bmatrix} \tag{16}$$

For notational convenience, we define $\Xi \in \mathbb{R}^{n\times n}$, $\Upsilon \in \mathbb{R}^n$ and $\Omega \in \mathbb{R}$ as

$$\begin{aligned}\Xi(\tau,N) &\triangleq \frac{1}{2N_\tau}\Psi_{0/\tau}^{\mathrm{T}}\Psi_{\tau/0} = \frac{1}{2N_\tau}\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau} \\ &= \frac{1}{2}\{\bar{R}_{\psi\psi}(t,t-\tau,N) + \bar{R}_{\psi\psi}(t-\tau,t,N)\}\end{aligned} \tag{17}$$

$$\begin{aligned}\Upsilon(\tau,N) &\triangleq \frac{1}{2N_\tau}\Psi_{0/\tau}^{\mathrm{T}}Y_{\tau/0} = \frac{1}{2N_\tau}\Psi_{0/\tau}^{\mathrm{T}}Y_{\tau/0} \\ &= \frac{1}{2}\{\bar{r}_{\psi y}(t,t-\tau,N) + \bar{r}_{\psi y}(t-\tau,t,N)\}\end{aligned} \tag{18}$$

Copyright © 2006 John Wiley & Sons, Ltd.

$$\Omega(\tau, N) \triangleq \frac{1}{2N_\tau} Y_{0/\tau}^{\mathrm{T}} Y_{\tau/0} = \frac{1}{2N_\tau} Y_{0/\tau}^{\mathrm{T}} Y_{\tau/0}$$
$$= \frac{1}{2}\{\bar{r}_{yy}(t, t-\tau, N) + \bar{r}_{yy}(t-\tau, t, N)\} \tag{19}$$

where $\bar{r}_{\psi y}(t_1, t_2, N)$ and $\bar{r}_{yy}(t_1, t_2, N)$ are empirical correlations defined as in (10). Minimizing (13) yields the least-correlation estimate.

Theorem 2 (*least-correlation estimate*)
Consider $J^2(\bar{\theta}, \tau, N)$ in (12) or (13). Suppose that A4 is satisfied. Then $\Xi(\tau, N)$ is nonsingular. Furthermore, $J^2(\bar{\theta}, \tau, N)$ has a unique minimum at

$$\hat{\theta}(\tau, N) = (\Psi_{0/\tau}^{\mathrm{T}} \Psi_{\tau/0})^{-1} \Psi_{0/\tau}^{\mathrm{T}} Y_{\tau/0} = \Xi^{-1}(\tau, N)\Upsilon(\tau, N) \tag{20}$$

and the corresponding minimum of $J^2(\hat{\theta}, \tau, N)$ is

$$J^2(\hat{\theta}(\tau, N), \tau, N) = \{\Omega(\tau, N) - \Upsilon^{\mathrm{T}}(\tau, N)\Xi^{-1}(\tau, N)\Upsilon(\tau, N)\}^2 \tag{21}$$

Proof
The matrix $\Xi(\tau, N)$ is nonsingular since the empirical correlations in (17) have full rank due to A4. Taking the gradient of (13) with respect to $\bar{\theta}$ yields the necessary condition

$$\Psi_{0/\tau}^{\mathrm{T}} \Psi_{\tau/0} \hat{\theta} = \Psi_{0/\tau}^{\mathrm{T}} Y_{\tau/0} \tag{22}$$

that is

$$\Xi(\tau, N)\hat{\theta} = \Upsilon(\tau, N) \tag{23}$$

where $\hat{\theta}$ minimizes (13). Adding and subtracting $\Upsilon^{\mathrm{T}}\Xi^{-1}\Upsilon$ in (13) and completing the squares yields

$$J^2(\bar{\theta}, \tau, N) = \{(\bar{\theta} - \Xi^{-1}\Upsilon)^{\mathrm{T}}\Xi(\bar{\theta} - \Xi^{-1}\Upsilon) + (\Omega - \Upsilon^{\mathrm{T}}\Xi^{-1}\Upsilon)\}^2 \tag{24}$$

Since the second term of (24) is independent of $\bar{\theta}$, (20) is unique and minimizes $J^2(\bar{\theta}, \tau, N)$ since A4 is satisfied. Substituting (20) into (24) yields (21). □

Remark
It is noted that (22) plays an important role similar to the normal equation of least squares. Setting $\tau = 0$ to (22) and (20) yields the normal equation

$$\Xi(0, N)\hat{\theta} = \Upsilon(0, N) \tag{25}$$

that is

$$\bar{R}_{\psi\psi}(0, N)\hat{\theta} = \bar{r}_{\psi y}(0, N) \tag{26}$$

and the least-squares estimate

$$\hat{\theta}(0, N) = \Xi^{-1}(0, N)\Upsilon(0, N) = \bar{R}_{\psi\psi}^{-1}(0, N)\bar{r}_{\psi y}(0, N) \tag{27}$$

respectively.

Now let us consider the convergence of the least-correlation estimate. As the number of samples tends to infinity, the empirical correlations in (17)–(19) converge to the corresponding

Copyright © 2006 John Wiley & Sons, Ltd.

mathematical correlations with probability 1 as given by

$$\lim_{N\to\infty} \Xi(\tau, N) = \lim_{N\to\infty} \frac{1}{2}\{\bar{R}_{\psi\psi}(t, t-\tau, N) + \bar{R}_{\psi\psi}(t-\tau, t, N)\} = R_{\psi\psi}(\tau) \tag{28}$$

$$\lim_{N\to\infty} \Upsilon(\tau, N) = r_{\psi y}(\tau) \tag{29}$$

$$\lim_{N\to\infty} \Omega(\tau, N) = r_{yy}(\tau) \tag{30}$$

due to A1–A3 and the ergodicity [13, Theorem 2.3, p. 43].

Theorem 3 (*consistency*)
Suppose that A1–A4 are satisfied. Let $\hat{\theta}(\tau, N)$ be defined by (20). Then

$$\lim_{N\to\infty} \hat{\theta}(\tau, N) = \theta \quad \text{with probability 1} \tag{31}$$

Proof
According to the ergodicity as well as A1–A2,

$$\sup_{\bar{\theta}} |J^2(\bar{\theta}, \tau, N) - \bar{J^2}(\bar{\theta}, \tau)| \to 0, \quad \text{with probability 1 as } N \to \infty \tag{32}$$

where $\bar{J^2}(\bar{\theta}, \tau) \triangleq \bar{E}^2[\varepsilon(t, \bar{\theta})\varepsilon(t-\tau, \bar{\theta})]$. The criterion $J^2(\bar{\theta}, \tau, N)$, therefore, converges uniformly to $\bar{J^2}(\bar{\theta}, \tau)$. This implies that the unique minimizing argument $\hat{\theta}(\tau, N)$ of J^2 converges to the minimizing argument $\hat{\theta}(\tau)$ of $\bar{J^2}$. We show the uniqueness of $\hat{\theta}(\tau)$ and $\hat{\theta}(\tau) = \theta$ as follows. When N goes to infinity, (22) and (24) gives the unique minimum

$$\hat{\theta}(\tau) = R_{\psi\psi}^{-1}(\tau) r_{\psi y}(\tau) \tag{33}$$

of $\bar{J^2}$ as in the proof of Theorem 2. Moreover, $\hat{\theta}(\tau)$ is evaluated as

$$\hat{\theta}(\tau) = R_{\psi\psi}^{-1}(\tau)\{R_{\psi\psi}(\tau)\theta + r_{\psi\eta}(\tau) - R_{\psi\zeta}(\tau)\theta\} = \theta \tag{34}$$

since $r_{\psi\eta}(\tau) = 0$, $R_{\psi\zeta}(\tau) = 0$ due to A3. □

Setting $\tau = 0$ in Theorem 3 gives the convergence property of least-squares which is equivalent to References [13, Theorem 8.2, p. 254], [14], [15, Theorem 5.2.1, pp. 226–227]. For the EIV model (1)–(5), the least-squares estimates $\hat{\theta}(0, N)$ in (27) generates the error-prone results as

$$\lim_{N\to\infty} \tilde{\theta}(0, N) = R_{\psi\psi}^{-1}(0)\{r_{\psi\eta}(0) - R_{\psi\zeta}(0)\theta\} \tag{35}$$

where $\tilde{\theta}(\tau, N) \triangleq \hat{\theta}(\tau, N) - \theta$, since $R_{\psi\zeta}(0)$ cannot be zero even if $\zeta(t)$ is white. The conventional least-squares estimate, therefore, is not consistent for EIV models.

Copyright © 2006 John Wiley & Sons, Ltd.

4. GEOMETRICAL ASPECT OF THE LEAST-CORRELATION ESTIMATE

Consider (22). Two equations

$$\Psi_{0/\tau}^{\mathrm{T}} Y_{\tau/0} - \Psi_{0/\tau}^{\mathrm{T}} \Psi_{\tau/0} \hat{\theta}(\tau) = 0 \tag{36}$$

$$\Psi_{\tau/0}^{\mathrm{T}} Y_{0/\tau} - \Psi_{\tau/0}^{\mathrm{T}} \Psi_{0/\tau} \hat{\theta}(\tau) = 0 \tag{37}$$

are equivalent to. Let us rewrite the estimated outputs as

$$\hat{Y}_{\tau/0}(\hat{\theta}(\tau)) = \Psi_{\tau/0} \hat{\theta}(\tau) \tag{38}$$

$$\hat{Y}_{0/\tau}(\hat{\theta}(\tau)) = \Psi_{0/\tau} \hat{\theta}(\tau) \tag{39}$$

and the corresponding residuals as

$$\mathscr{E}_{\tau/0}(\hat{\theta}(\tau)) = Y_{\tau/0} - \Psi_{\tau/0} \hat{\theta}(\tau) \tag{40}$$

$$\mathscr{E}_{0/\tau}(\hat{\theta}(\tau)) = Y_{0/\tau} - \Psi_{0/\tau} \hat{\theta}(\tau) \tag{41}$$

From the above expressions, we obtain the following property:

Lemma 4 (*orthogonality*)
Suppose that A4 is satisfied. Let $\hat{\theta}(\tau)$ be defined by (20). Then

$$\Psi_{0/\tau}^{\mathrm{T}} \mathscr{E}_{\tau/0} = \Psi_{\tau/0}^{\mathrm{T}} \mathscr{E}_{0/\tau} = 0 \tag{42}$$

Proof
Using (40), (41)–(36), (37) yields (42). □

Corollary 5
Property (42) is equivalent to

$$\hat{Y}_{0/\tau}^{\mathrm{T}} \mathscr{E}_{\tau/0} = \hat{Y}_{\tau/0}^{\mathrm{T}} \mathscr{E}_{0/\tau} = 0 \tag{43}$$

Proof
Left-multiplying (42) by $\hat{\theta}^{\mathrm{T}}(\tau)$ and employing (38), (39) gives (43). □

Let $\mathbf{Y}_{\tau/0}$ and $\mathbf{Y}_{0/\tau}$ be linear vector spaces, $Y_{\tau/0} \in \mathbf{Y}_{\tau/0}$ and $Y_{0/\tau} \in \mathbf{Y}_{0/\tau}$, respectively, and let $\hat{\mathbf{Y}}_{\tau/0}$ and $\hat{\mathbf{Y}}_{0/\tau}$ be corresponding subspaces, $\hat{Y}_{\tau/0} \in \hat{\mathbf{Y}}_{\tau/0} \subset \mathbf{Y}_{\tau/0}$ and $\hat{Y}_{0/\tau} \in \hat{\mathbf{Y}}_{0/\tau} \subset \mathbf{Y}_{0/\tau}$. Let $P_{\tau/0}$ and $P_{0/\tau}$, defined by

$$P_{\tau/0} \triangleq \Psi_{0/\tau} \{\Psi_{\tau/0}^{\mathrm{T}} \Psi_{0/\tau}\}^{-1} \Psi_{\tau/0}^{\mathrm{T}} \tag{44}$$

$$P_{0/\tau} \triangleq \Psi_{\tau/0} \{\Psi_{0/\tau}^{\mathrm{T}} \Psi_{\tau/0}\}^{-1} \Psi_{0/\tau}^{\mathrm{T}} \tag{45}$$

Copyright © 2006 John Wiley & Sons, Ltd.

operate on $\mathbf{Y}_{\tau/0}$ and $\mathbf{Y}_{0/\tau}$, respectively. Then the operators $P_{\tau/0}$, $P_{0/\tau}$ map the measurements $Y_{\tau/0}$, $Y_{0/\tau}$ to the estimates $\hat{Y}_{\tau/0}$, $\hat{Y}_{0/\tau}$ as

$$\hat{Y}_{\tau/0}(\hat{\theta}(\tau)) = P_{\tau/0} Y_{\tau/0} \tag{46}$$

$$\hat{Y}_{0/\tau}(\hat{\theta}(\tau)) = P_{0/\tau} Y_{0/\tau} \tag{47}$$

With an identity matrix $I \in \mathbb{R}^{N_\tau \times N_\tau}$, let us define a matrix

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \tag{48}$$

Note that $J^2 = I$. Observations on $P_{\tau/0}$ and $P_{0/\tau}$ include the following properties summarized in Figure 1 and Lemma 6. Figure 1 depicts the relationships among the operators $P_{\tau/0}$, $P_{0/\tau}$, the measurements $Y_{\tau/0}$, $Y_{0/\tau}$, $\Psi_{\tau/0}$, $\Psi_{0/\tau}$, the estimates $\hat{Y}_{\tau/0}$, $\hat{Y}_{0/\tau}$, and their residuals $\mathscr{E}_{\tau/0}$, $\mathscr{E}_{0/\tau}$.

Lemma 6
$P_{\tau/0}$ and $P_{0/\tau}$ have the following properties:

1. $P_{\tau/0}$ and $P_{0/\tau}$ are oblique projection operators.
2. $P_{\tau/0} = P_{0/\tau}^{\mathrm{T}}$ and $P_{0/\tau} = P_{\tau/0}^{\mathrm{T}}$.
3. $P_{\tau/0} = JP_{0/\tau}J$ and $P_{0/\tau} = JP_{\tau/0}J$.
4. $(I - P_{0/\tau})J = J(I - P_{\tau/0})$ and $(I - P_{\tau/0})J = J(I - P_{0/\tau})$.

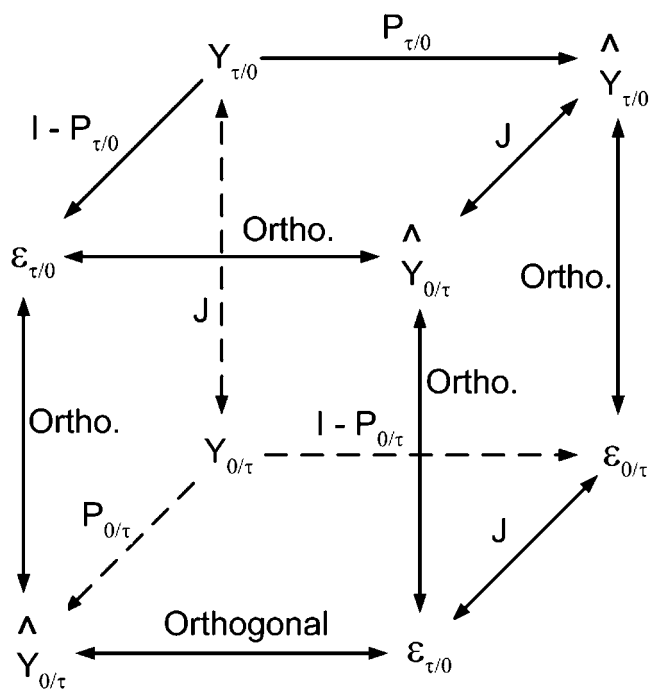

Figure 1. Relationship among operators, measurements, estimates, and/or residuals.

Copyright © 2006 John Wiley & Sons, Ltd.

5. Both $(I - P_{0/\tau})J$ and $(I - P_{\tau/0})J$ are self-adjoint.
6. Suppose that $J(I - P_{\tau/0})$ and $J(I - P_{0/\tau})$ operate on $\mathbf{Y}_{\tau/0}$ and $\mathbf{Y}_{0/\tau}$, respectively. Then range$[P_{\tau/0}]$ is orthogonal to range$[J(I - P_{\tau/0})]$, and range$[P_{0/\tau}]$ is orthogonal to range $[J(I - P_{0/\tau})]$.

Proof
Each item is composed of two statements. It is enough to sketch one of them.

1. $P_{\tau/0}$ is idempotent, $P_{\tau/0}^2 = P_{\tau/0}$, but not self-adjoint, $P_{\tau/0} \neq P_{\tau/0}^{\mathrm{T}}$ [16, 17, p. 71].
2. The statement follows immediately from (44) and (45).
3. Applying the equivalent expression

$$P_{0/\tau} = \begin{bmatrix} \Psi_\tau\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_0^{\mathrm{T}} & \Psi_\tau\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_\tau^{\mathrm{T}} \\ \Psi_0\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_0^{\mathrm{T}} & \Psi_0\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_\tau^{\mathrm{T}} \end{bmatrix} \tag{49}$$

to the right-hand side of the first equality shows

$$\begin{aligned} JP_{0/\tau}J &= \begin{bmatrix} \Psi_0\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_\tau^{\mathrm{T}} & \Psi_0\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_0^{\mathrm{T}} \\ \Psi_\tau\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_\tau^{\mathrm{T}} & \Psi_\tau\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_0^{\mathrm{T}} \end{bmatrix} \\ &= \Psi_{0/\tau}\{\Psi_{\tau/0}^{\mathrm{T}}\Psi_{0/\tau}\}^{-1}\Psi_{\tau/0}^{\mathrm{T}} = P_{\tau/0} \end{aligned} \tag{50}$$

4. $J(I - P_{\tau/0}) = J - JP_{\tau/0} = J - JP_{\tau/0}JJ = J - P_{0/\tau}J = (I - P_{0/\tau})J$.
5. Since we confine our discussion to real data in this work, it is enough to show the self-symmetry instead of the self-adjoint as

$$[(I - P_{0/\tau})J]^{\mathrm{T}} = J - JP_{\tau/0}JJ = J - P_{0/\tau}J = (I - P_{0/\tau})J \tag{51}$$

6. Consider an arbitrary vector $Y_{\tau/0} \in \mathbf{Y}_{\tau/0}$. Applying $P_{\tau/0}$ to $Y_{\tau/0}$ generates $\hat{Y}_{\tau/0}$ from (46) and using $(I - P_{0/\tau})J$ to $Y_{\tau/0}$ yields

$$(I - P_{0/\tau})JY_{\tau/0} = (I - P_{0/\tau})Y_{0/\tau} = Y_{0/\tau} - \hat{Y}_{0/\tau} = \mathscr{E}_{0/\tau} \tag{52}$$

According to Corollary 5, $\mathscr{E}_{0/\tau}$ and $\hat{Y}_{\tau/0}$ are orthogonal. Thus the spaces spanned by $\mathscr{E}_{0/\tau}$ and $\hat{Y}_{\tau/0}$ are also orthogonal. □

5. NUMERICAL EXAMPLE

Fourier series is an optimal fitting in the sense of least squares [18]. We modify the problem into the EIV setting with

$$\phi(t) = [\sin 2\pi t \;\; \sin 6\pi t]^{\mathrm{T}}$$

$$\theta = [4/\pi \;\; 4/3\pi]^{\mathrm{T}}$$

Copyright © 2006 John Wiley & Sons, Ltd.

$$z(t) = \text{sign}[\sin 2\pi t]$$

$$\eta_1(t) = z(t) - \phi^{\mathrm{T}}(t)\theta$$

$$\eta_2(t) = \frac{0.2}{1 - 0.8q^{-1}}\xi_\eta(t)$$

$$\zeta_i(t) = G_\zeta(q^{-1})\xi_{\zeta_i}(t), \quad i = 1, 2$$

where $\xi_\eta(t)$ and $\xi_{\zeta_i}(t)$ are white with variances chosen such that the signal-to-noise ratios

$$\text{SNR}_\text{i} = 10\log_{10}\left(\frac{\bar{E}[\phi^{\mathrm{T}}(t)\phi(t)]}{E[\zeta^{\mathrm{T}}(t)\zeta(t)]}\right) \tag{53}$$

$$\text{SNR}_\text{o} = 10\log_{10}\left(\frac{\bar{E}[z^2(t)]}{E[\eta^2(t)]}\right) \tag{54}$$

are about 20, 15, 10, 5, or 0 dB, but the maximum SNR_o is limited to about 10 dB due to the modelling residual $\eta_1(t)$. According to A3, the regressor noise $\zeta(t)$ should be at most finitely correlated, but in this example we consider an infinitely correlated case as well as a finitely correlated noise as follows:

Case 1:

$$\text{FIR}: \quad G_\zeta(q^{-1}) = 0.3 + 0.7q^{-1} \tag{55}$$

Case 2:

$$\text{IIR}: \quad G_\zeta(q^{-1}) = \frac{0.3}{1 - 0.7q^{-1}} \tag{56}$$

Figures 2–4 and Tables I and II summarize the simulation results from 100 Monte Carlo runs for each case and some theoretical evaluations. In each figure '+' denotes the true parameter

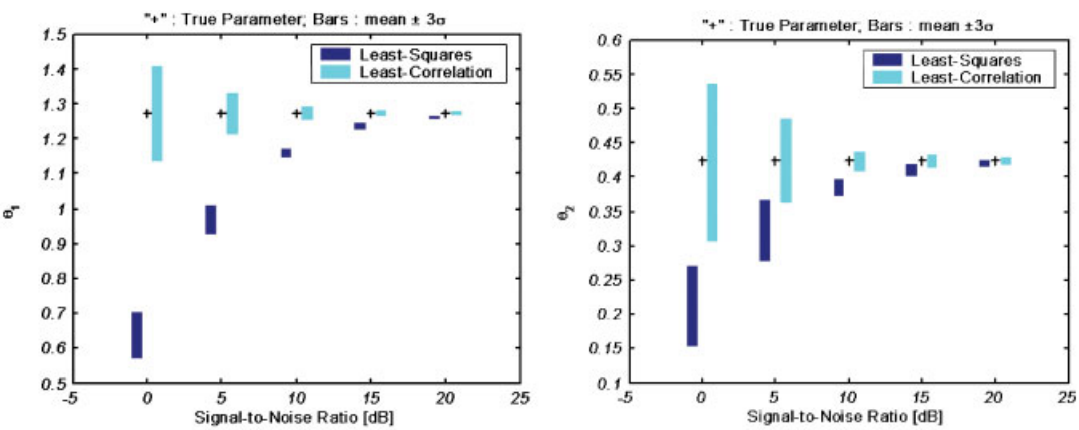

Figure 2. Effect of SNR: FIR input noise, 10^4 samples.

Copyright © 2006 John Wiley & Sons, Ltd.

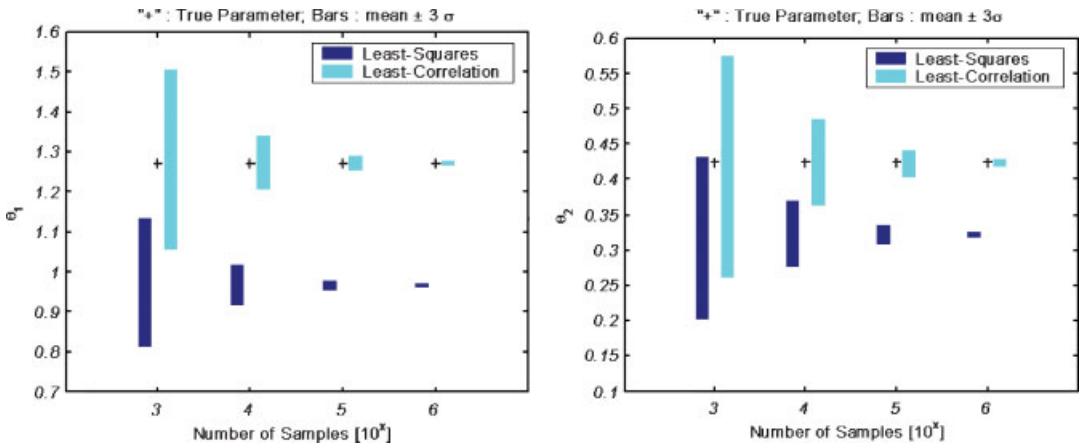

Figure 3. Effect of the number of samples: FIR input noise, $\mathrm{SNR_i} \simeq \mathrm{SNR_o} \simeq 5$ dB.

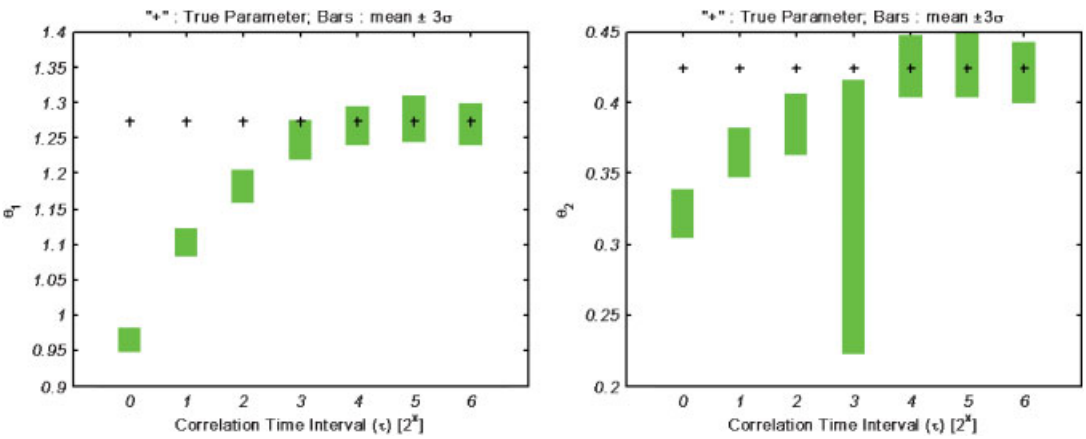

Figure 4. Effect of τ on IIR input noise: $\mathrm{SNR_i} \simeq \mathrm{SNR_o} \simeq 5$ dB, 10^4 samples.

Table I. Comparison of empirical estimates and theoretical calculations: FIR input noise, 10^4 samples.

Estimate	$\mathrm{SNR_i}$ (dB)	$\mathrm{SNR_o}$ (dB)	Theory (%) $E[\tilde{\theta}_1]$	Theory (%) $E[\tilde{\theta}_2]$	Simulation (%) $\bar{\tilde{\theta}}_1$	Simulation (%) $\bar{\tilde{\theta}}_2$
	20	10	−1.0	−1.3	−1.02	−1.31
	15	10	−3.1	−3.3	−3.07	−3.46
$\hat{\theta}(0)$	10	10	−9.0	−9.3	−9.06	−9.32
	5	5	−24.0	−24.2	−24.07	−24.05
	0	0	−49.9	−50.0	−50.00	−50.07
	20	10	0.0	0.0	−0.04	−0.31
	15	15	0.0	0.0	−0.01	−0.42
$\hat{\theta}(\tau)$	10	10	0.0	0.0	−0.08	−0.36
$(\tau = 2)$	5	5	0.0	0.0	−0.07	−0.08
	0	0	0.0	0.0	−0.18	−0.94

Copyright © 2006 John Wiley & Sons, Ltd.

Table II. Comparison of empirical estimates and theoretical calculations: IIR input noise, $\mathrm{SNR_i} \simeq \mathrm{SNR_o} \simeq 5$ dB, 10^4 samples.

Estimate	τ	Theory (%) $E[\tilde{\theta}_1]$	Theory (%) $E[\tilde{\theta}_2]$	Simulation (%) $\bar{\tilde{\theta}}_1$	Simulation (%) $\bar{\tilde{\theta}}_2$
$\hat{\theta}(\tau)$	0	−24.0	−24.2	−24.22	−24.33
	2	−13.5	−14.3	−13.41	−14.03
	4	−7.3	−9.4	−7.19	−9.45
	8	−2.0	−23.3	−2.04	−24.83
	16	−0.2	0.1	−0.48	0.27
	32	0.0	0.0	0.32	0.43
	64	0.0	0.0	−0.36	−0.72

and the rectangular means the expected range of corresponding estimate defined by

$$\hat{\hat{\theta}}_i = \bar{\hat{\theta}}_i \pm 3\bar{\sigma}(\hat{\theta}_i), \quad i = 1, 2 \tag{57}$$

where $\bar{\hat{\theta}}_i$ and $\bar{\sigma}(\hat{\theta}_i)$ denote the empirical mean and the empirical standard deviation of the estimate $\hat{\theta}_i$, respectively. Tables I and II compare theoretical errors $E[\tilde{\theta}_i]$, which are evaluated by the corresponding equations (33) or (35), with the empirical errors $\bar{\tilde{\theta}}_i$ from simulations. For the IIR noise case in Table II, the correlations with sufficiently large time intervals are counted in. Generally speaking based on Tables I and II, there are good agreements between the theoretical calculations and the Monte Carlo simulations.

Before proceeding, let us sketch why the case with $\tau = 8$ in Table II and Figure 4 goes against the trend that the identification error decreases as τ increases. Consider the empirical correlation matrix

$$\bar{R}_{\phi\phi}(t, t-\tau, N) \triangleq \frac{1}{N_\tau} \sum_{t=1}^{N-\tau} \phi(t)\phi^{\mathrm{T}}(t+\tau)$$

which is a deterministic component of $\Xi(\tau, N)$. In this example, $\bar{R}_{\phi\phi}(t, t-\tau, N)$ is singular if $\tau h = k/12$, $k = \pm 1, \pm 3, \pm 5, \ldots$, where $h = 0.01$ in this example. Analysis of the singularity shows that one of two eigenvalues of $\bar{R}_{\phi\phi}(t, t-\tau, N)$ is zero if $\tau h = \pm k/12$ and $\tau h \neq \pm k/4$. The case $\tau = 8$ is sufficiently close to the singularity condition $\tau h = \pm 1/12$ that the least-correlation algorithm does not work well on one out of two parameters.

Figures 2 and 3 confirm numerically that the method of least correlation works well for the EIV problem. Figure 3 states that the least-correlation estimate can provide almost exact parameters from severely contaminated data provided that the number of samples is sufficiently large. We try to show in Figure 4 that the method of least correlation can be applied to the problems with regressors corrupted by infinitely correlated noise which apparently violates A3. Figure 4 says that the least-correlation method can give useful estimates if A4 is satisfied with sufficiently large τ. Table II shows that the simulation results in Figure 4 are also supported by the theoretical calculations.

Copyright © 2006 John Wiley & Sons, Ltd.

6. RLC ALGORITHMS

With the augmented regressors $\psi_{i/i-\tau} \in \mathbb{R}^{n\times 2}$, $\psi_{i-\tau/i} \in \mathbb{R}^{n\times 2}$ and the augmented output $y_{i-\tau/i} \in \mathbb{R}^2$ defined by

$$\psi_{i/i-\tau} \triangleq [\psi(i) \;\; \psi(i-\tau)]$$

$$\psi_{i-\tau/i} \triangleq [\psi(i-\tau) \;\; \psi(i)]$$

$$y_{i-\tau/i} \triangleq [y(i-\tau) \;\; y(i)]^{\mathrm{T}}$$

an equivalent expression of (20) is written as

$$\hat{\theta}(\tau, t) = \left(\sum_{i=1+\tau}^{t} \psi_{i/i-\tau}\psi_{i-\tau/i}^{\mathrm{T}} \right)^{-1} \sum_{i=1+\tau}^{t} \psi_{i/i-\tau} y_{i-\tau/i} \tag{58}$$

Employing the steps given in Reference [19, pp. 262–263], which derives the recursive least-squares (RLS) algorithm from its off-line version—the least-squares estimate (27), for the least-correlation estimate (20) yields the RLC algorithm

$$\hat{\theta}(\tau, t) = \hat{\theta}(\tau, t-1) + K(t)(y_{t-\tau/t} - \psi_{t-\tau/t}^{\mathrm{T}}\hat{\theta}(\tau, t-1)) \tag{59}$$

$$K(t) = P(t-1)\psi_{t/t-\tau}(I + \psi_{t-\tau/t}^{\mathrm{T}}P(t-1)\psi_{t/t-\tau})^{-1} \tag{60}$$

$$P(t) = P(t-1) - K(t)\psi_{t-\tau/t}^{\mathrm{T}}P(t-1) \tag{61}$$

for $t > \tau$ provided that the initial values $P(\tau)$ and $\hat{\theta}(\tau, \tau)$ are given. The recursive algorithm (59)–(61) is equivalent to (20), which means that the estimates from both algorithms are the same at the final time. Figure 5 shows the estimates for the example in Section 5 by the RLC algorithm (59)–(61) and the RLS algorithm [13, 19].

On the other hand, for the continuous-time expression of (12)

$$J_c^2(\bar{\theta}, \tau, t) = \left(\frac{1}{t-\tau} \int_{\tau}^{t} \varepsilon(s, \bar{\theta})\varepsilon(s-\tau, \bar{\theta})\, \mathrm{d}s \right)^2 \tag{62}$$

applying the minimization procedure [20, pp. 370–371] gives the continuous-time RLC algorithm

$$\dot{P}(t) = -P(t)[\psi(t)\psi^{\mathrm{T}}(t-\tau) + \psi(t-\tau)\psi^{\mathrm{T}}(t)]P(t) \tag{63}$$

$$\dot{\hat{\theta}}(t) = -P(t)[\psi(t)e(t-\tau|t) + \psi(t-\tau)e(t|t)] \tag{64}$$

for $t > \tau$, where the errors $e(t-\tau|t)$ and $e(t|t)$ are defined by

$$e(t-\tau|t) = \psi^{\mathrm{T}}(t-\tau)\hat{\theta}(t) - y(t-\tau) \tag{65}$$

$$e(t|t) = \psi^{\mathrm{T}}(t)\hat{\theta}(t) - y(t) \tag{66}$$

respectively. Note that t in (62)–(66) denotes the time in the continuous-time domain.

Copyright © 2006 John Wiley & Sons, Ltd.

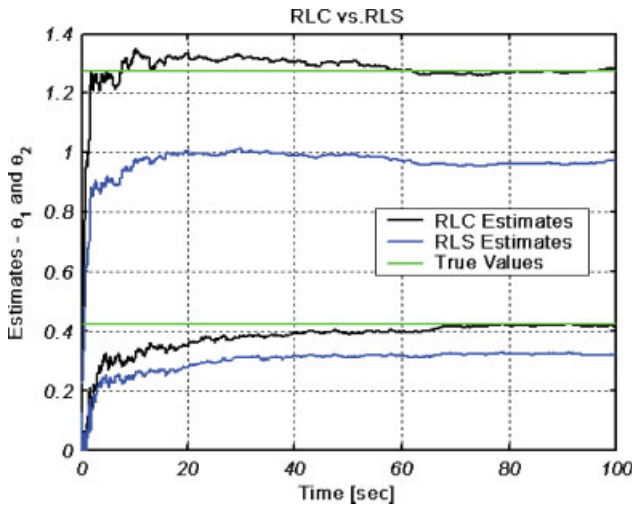

Figure 5. Online estimates by RLS and RLC algorithms, $\theta_1 = 4/\pi$, $\theta_2 = 4/3\pi$: FIR input noise, $\tau = 2$, $\mathrm{SNR_i} = \mathrm{SNR_o} = 5$ dB, sampling 10 ms.

7. CONCLUDING REMARKS

Based on observations about the residuals from least-squares applied to EIV models, we introduce a criterion defined by squaring empirical correlation between residuals. The necessary and sufficient condition minimizing the criterion yields a new identification procedure called the least-correlation estimate. The estimate generalizes the well-known least-squares since the least-correlation becomes the least-squares when the correlation lag goes to zero. We show that the estimate converges to the true value as the number of samples tends to infinity. Monte Carlo simulations also support the consistency numerically. Moreover, the numerical results suggest that the estimate can deal with the infinitely correlated noise as well as the finitely correlated noise. Analysis shows that the least-correlation estimate has an interesting geometrical property which is partly similar to that of the least-squares estimate. One of the differences is that the least-correlation estimate is not an orthogonal projection but an oblique projection. Finally, we derive the RLC algorithm on both the continuous-time and discrete-time domains. A simulation demonstrates that the recursive version works well.

It is possible to interpret the least-correlation estimate as a deterministic representation of the 'error whitening Wiener filter' [11]. The previous works [11, 12] provide expressions for unbiasedness and orthogonality. In this paper, we also discussed sufficiency for a minimum, the consistency of the estimate, the geometrical interpretation of the estimator, and the recursive realization on continuous-time domain.

This work is extended to the estimates for EIV nonlinear models [21]. Input design problem, which is closely related to A4, for the best estimate is a further work for application. Another study will be a kind of realization problem for the RLC algorithm to be numerically robust and computationally efficient.

Copyright © 2006 John Wiley & Sons, Ltd.

REFERENCES

1. Zarchan P, Musoff H. *Fundamentals of Kalman Filtering*: *A Practical Approach*. AIAA: New York, 2000.
2. Feng CB, Zheng WX. Robust identification of stochastic linear systems with correlated noise. *IEE Proceedings*, *Part D* 1991; **138**(5):484–492.
3. van Huffel S, Vandewalle J. *The Total Least Squares Problem*: *Computational Aspects and Analysis*. SIAM: Philadelphia, PA, 1991.
4. Frish R. Statistical confluence analysis by means of complete regression systems. *Technical Report 5*, University of Oslo, Economic Institute, Oslo, Norway, 1934.
5. Fuller WA. *Measurement Error Models*. Wiley: Berlin, 1987.
6. Koopmans TJ. *Linear Regression Analysis of Economic Time Series*. De Erven F. Bohn: N. V. Haarlem, The Netherlands, 1937.
7. Levin ML. Estimation of a system pulse transfer function in the presence of noises. *IEEE Transactions on Automatic Control* 1964; **9**:229–235.
8. Scherrer W, Deistler M. A structure theory for linear dynamic errors-in-variables models. *SIAM Journal on Control Optimization* 1998; **36**(6):2148–2175.
9. Soderstrom T, Soverini U, Mahata K. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing* 2002; **82**:1139–1154.
10. van Huffel S, Lemmerling P (eds). *Total Least Squares Techniques and Errors-in-Variables Modeling*: *Analysis*, *Algorithms and Applications*. Kluwer Academic Publishers: Dordrecht, 2002.
11. Principe JC, Rao YN, Erdogmus D. Error whitening Wiener filters: theory and algorithms. In *Advances in LMS Filters*, Kaykin S, Widrow B (eds). Wiley: New York, 2003.
12. Rao YN, Erdogmus D, Rao GY, Principe JC. Stochastic error whitening algorithm for linear filter estimation with noisy data. *Neural Networks* 2003; **16**:873–990.
13. Ljung L. *System Identification—Theory for the User* (2nd edn). Prentice-Hall: Englewood Cliffs, NJ, 1999.
14. Zheng WX. Transfer function estimation from noisy input and output data. *International Journal of Adaptive Control and Signal Processing* 1998; **12**:365–380.
15. Davis MHA, Vinter RB. *Stochastic Modelling and Control*. Chapman & Hall: London, 1985.
16. Behrens RT, Scharf LL. Signal processing applications of oblique projection operators. *IEEE Transactions on Signal Processing* 1994; **42**(6):1413–1424.
17. Fuhrmann PA. *Linear Systems and Operators in Hilbert Space*. McGraw-Hill: New York, 1981.
18. Bay JS. *Fundamentals of Linear Space Systems*. McGraw-Hill: New York, 1999.
19. Johansson R. *System Modeling and Identification*. Prentice-Hall: Englewood Cliffs, NJ, 1993.
20. Slotine JE, Li W. *Applied Nonlinear Control*. Prentice-Hall: Englewood Cliffs, NJ, 1991.
21. Jun BE, Bernstein DS. Least-correlations estimates for errors-in-variables nonlinear models. *IEEE Industrial Electronics Conference*, FC6-4, 2004.

Copyright © 2006 John Wiley & Sons, Ltd.