Least-Correlation Estimates for Errors-in-Variables Models

Byung-Eul Jun[1] and Dennis S. Bernstein[2]

[1]Principal Researcher, Agency for Defense Development, Daejeon 305-600, Korea
[2]Professor, Aerospace Eng., University of Michigan, Ann Arbor, MI 48109, USA
email: [1]bejun@unitel.co.kr, [2]dsbaero@umich.edu

Abstract—This paper introduces an estimator working on errors-in-variables models whose all variables are corrupted by noise. The necessary and sufficient condition minimizing the criterion, defined by the square of empirical correlation between residuals with a non-zero time interval, gives the least-correlation estimates. The method of least correlation can be interpreted as a generalization of the least-squares. Analysis shows that the estimator has a capability to find out the best fit without bias from noisy measurements even contaminated by colored noise as the number of observations increases. Monte Carlo simulations for numerical examples support the consistency of the estimator. The least-correlation estimate is not an orthogonal projection but an oblique projection. We discuss interesting geometric properties of the estimate. Finally recursive realizations of the estimator in continuous-time domain as well as in discrete-time are mentioned briefly.

I. INTRODUCTION

The method of least squares is still the most popular approach to finding out the best fit to a given structure [20], but it exhibits high sensitivity to errors in regressors [3,19]. A generalized approach to modelling noise is to view all variables as contaminated by noise, called errors-in-variables (EIV) models [12,14,18] which has a broad application in time series modelling, image processing, signal processing, neural networks and system identification. A recent trend in systems is trying to use powerful computers with low grade instruments, which implies in general that the need for EIV models is increasing.

We introduce a criterion defined by the square of empirical autocorrelation between residuals with a non-zero time interval and derive an estimate minimizing the criterion. Analysis shows that the optimality in the sense of least correlation has several attractive features. It works on static systems as well as dynamic systems since it does not depend on any specific structure of regressors. Only simple matrix algebra gives the estimate. There is a direct relationship between the criterion and the stochastic correlation. It works on EIV models with colored noises as well as with white noises. Literature reveals that the 'error whitening Wiener filter' [10,11] can be considered as a stochastic counterpart of the least-correlation estimate introduced in this work. Under the setting in the previous works [10,11], the estimation residuals are whitened by the 'error whitening Wiener filter'. In the generalized formulation in this paper, however, the estimate can not whiten the residuals any more but minimize the magnitude of correlation. In this sense, the term 'error whitening' can not be used for this work.

This paper reports some aspects of the least-correlation estimate besides its consistency. The estimate is the sufficient and necessary condition to minimize the criterion. The method of least correlation yields four induced estimators

and one of them can be interpreted as the instrumental variable estimator [8,9]. Analysis and simulations show that the induced estimators also work on EIV models. The least-correlation estimate shows interesting geometric property that is partly similar to that of the least-squares estimates. The estimate is not an orthogonal projection but a kind of oblique projection.

II. PROBLEM FORMULATION

Consider the linear regression model

$$z(t) = \phi^T(t)\theta + \eta_1(t), \qquad (1)$$

where $z(t) \in \mathbb{R}$ is the system response at tth sampling, $\phi(t) \in \mathbb{R}^n$ is the regression vector, $\theta \in \mathbb{R}^n$ is the parameter vector to be estimated 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 FIR (finite impulse response) systems, delayed outputs in all-pole systems or in AR (autoregression) models, and both delayed inputs and delayed outputs in ARX (autoregression with exogenous variables) models.
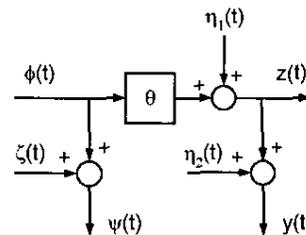

Fig. 1. Description of errors-in-variables linear regression models

Let $y(t)$ and $\psi(t)$ denote noise-corrupted measurements of $z(t)$ and $\phi(t)$, respectively, that is,

$$
\begin{aligned}
y(t) &= z(t) + \eta_2(t), & (2) \\
\psi(t) &= \phi(t) + \zeta(t), & (3)
\end{aligned}
$$

where $\eta_2(t) \in \mathbb{R}$ and $\zeta(t) \in \mathbb{R}^n$ are additive noises. Taking into account the noises included in both the regressors and the outputs constitutes an EIV problem [9,12,14,18], depicted in Fig. 1. In the approach in Fig. 1, the system (1) can be static or dynamic as far as it is written in linear regression model. Applying (2)-(3) to (1) yields

$$
\begin{aligned}
y(t) &= \psi^T(t)\theta + e(t), & (4) \\
e(t) &= \eta(t) - \zeta^T(t)\theta, & (5)
\end{aligned}
$$

where $\eta(t) \triangleq \eta_1(t) + \eta_2(t)$ denotes the total error on output. Now let us state the estimation problem for the EIV models.

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

Estimation problems frequently work with signals that are described as stochastic processes with deterministic components. For a common framework for deterministic and stochastic signals, 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 [9, p.34]. 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 asymptotically stable.

A2. The measured signals $\psi(t)$ and $y(t)$ are quasi-stationary and jointly quasi-stationary.

A3. The noises $\eta(t)$ and $\zeta(t)$ are zero-mean and finitely cross-correlated with $\psi(t)$, *i.e.*, there exists $\tau > 0$ such that

$$\bar{E}\left[\psi(t)\zeta^T(t-k)\right] = 0 \quad \text{for all } |k| \geq \tau, \tag{7}$$
$$\bar{E}\left[\psi(t)\eta(t-k)\right] = 0 \quad \text{for all } |k| \geq \tau. \tag{8}$$

A4. For τ in A3, $\psi(t)$ satisfies

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

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

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

Assumptions A3 and A4 express the idea that the correlations between data are stronger than those between noises themselves as well as those between signals and noises. Conditions (7) and (8) are equivalent to

$$\bar{E}\left[\phi(t)\zeta^T(t-k)\right] = 0, \quad \bar{E}\left[\zeta(t)\zeta^T(t-k)\right] = 0, \tag{11}$$
$$\bar{E}\left[\phi(t)\eta(t-k)\right] = 0, \quad \bar{E}\left[\zeta(t)\eta(t-k)\right] = 0, \tag{12}$$

for all $|k| \geq \tau > 0$, respectively, due to (3) and the arbitrariness of $\phi(t)$, $\zeta(t)$ and $\eta(t)$.

III. LEAST-CORRELATION ESTIMATES (LCE)

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

$$\epsilon(t, \bar{\theta}) = y(t) - \psi^T(t)\bar{\theta} \tag{13}$$

which models the mismatch between the observation on the real system and the behavior of the estimated model. If the sequence of residuals resulted from the least-squares estimate is not white, then at least one of the following statements is true.

1) Either the modelling or the estimate is not complete.
2) Either $\eta(t)$ or $\zeta(t)$ is colored.
3) There is a non-zero correlation between $\psi(t)$ and either $\eta(t)$ or $\zeta(t)$.

Applying (4)-(5) to (13) gives above condition 1) and 2). Condition 3) follows from the analysis of the least-squares estimate applied to the EIV problems (4)-(5) [3]. It is noted that the correlation between $\psi(t)$ and $\zeta(t)$ is not zero whenever $\zeta(t)$ is not zero, which is induced from (42) in Section IV.

Given $\bar{\theta} = \theta$ and $\eta(t) = 0$ in the EIV models, the correlation between $\epsilon(t, \bar{\theta})$ and $\epsilon(t', \bar{\theta})$, defined by $\bar{E}[\epsilon(t, \bar{\theta})\epsilon(t', \bar{\theta})]$, is zero due to A3 for all $|t - t'| \geq \tau$, but the mean square error $\bar{E}[\epsilon^2(t, \bar{\theta})]$ can never be zero. Based on this insight, we introduce a least-correlation criterion

$$J^2(\bar{\theta}, \tau, N) = \left(\frac{1}{N_\tau} \sum_{t=1+\tau}^{N} \epsilon(t, \bar{\theta})\epsilon(t-\tau, \bar{\theta})\right)^2, \tag{14}$$

where $N_{g(\tau)} \triangleq N - g(\tau)$, τ is an integer defining the time interval of autocorrelation and N denotes the number of samples. The cost function (14) can be rewritten as

$$J^2(\bar{\theta}, \tau, N) = \left(\frac{1}{N_\tau} \left(Y_0 - \Psi_0 \bar{\theta}\right)^T \left(Y_\tau - \Psi_\tau \bar{\theta}\right)\right)^2 \tag{15}$$

with the vectors Y_0, Y_τ and the matrices Ψ_0, Ψ_τ defined by

$$Y_0 \triangleq \begin{bmatrix} y(N) \\ y(N-1) \\ \vdots \\ y(1+\tau) \end{bmatrix}, \quad Y_\tau \triangleq \begin{bmatrix} y(N_\tau) \\ y(N_{\tau+1}) \\ \vdots \\ y(1) \end{bmatrix}, \tag{16}$$

$$\Psi_0 \triangleq \begin{bmatrix} \psi^T(N) \\ \psi^T(N-1) \\ \vdots \\ \psi^T(1+\tau) \end{bmatrix}, \quad \Psi_\tau \triangleq \begin{bmatrix} \psi^T(N_\tau) \\ \psi^T(N_{\tau+1}) \\ \vdots \\ \psi^T(1) \end{bmatrix}. \tag{17}$$

Minimizing (15) yields the estimate in Theorem 2. For notational simplicity in Theorem 2 and its proof we define

$$\Psi_{0/\tau} \triangleq \begin{bmatrix} \Psi_0 \\ \Psi_\tau \end{bmatrix}, \quad \Psi_{\tau/0} \triangleq \begin{bmatrix} \Psi_\tau \\ \Psi_0 \end{bmatrix}, \tag{18}$$

$$Y_{\tau/0} \triangleq \begin{bmatrix} Y_\tau \\ Y_0 \end{bmatrix}, \quad Y_{0/\tau} \triangleq \begin{bmatrix} Y_0 \\ Y_\tau \end{bmatrix}. \tag{19}$$

Theorem 2. (*Least-Correlation Estimate*) Consider the cost function $J^2(\bar{\theta}, \tau, N)$ given by (14) or (15). Suppose that A4 is satisfied. Then $\Psi_{0/\tau}^T \Psi_{\tau/0}$ is nonsingular. Furthermore $J^2(\bar{\theta}, \tau, N)$ has a unique minimum at

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

and the corresponding minimum is given by

$$J^2 = \left(\frac{1}{2N_\tau} Y_{0/\tau}^T \left\{I - \Psi_{\tau/0} \Xi^{-1} \Psi_{0/\tau}^T\right\} Y_{\tau/0}\right)^2 \tag{21}$$

with

$$\Xi \triangleq \Psi_{0/\tau}^T \Psi_{\tau/0} = \Psi_{\tau/0}^T \Psi_{0/\tau}. \tag{22}$$

Proof. Taking the gradient of (15) with respect to $\bar{\theta}$ yields a necessary condition

$$\Psi_{0/\tau}^T \Psi_{\tau/0} \hat{\theta} = \Psi_{0/\tau}^T Y_{\tau/0} \tag{23}$$

minimizing (15), where $\hat{\theta}$ denotes an optimal estimate. We can rewrite (15) by using (18)-(19) as

$$J^2 = \left\{\frac{1}{2N_\tau} \left(Y_{0/\tau} - \Psi_{0/\tau}\bar{\theta}\right)^T \left(Y_{\tau/0} - \Psi_{\tau/0}\bar{\theta}\right)\right\}^2. \tag{24}$$

Adding and subtracting

$$Y_{0/\tau}^T \Psi_{\tau/0} \left(\Psi_{0/\tau}^T \Psi_{\tau/0} \right)^{-1} \Psi_{0/\tau}^T Y_{\tau/0}$$

on the right-hand of (24) and completing the squares yields

$$J^2 = \left(\frac{1}{2N_\tau} \left\{ \bar{\theta} - \Xi^{-1}\Upsilon \right\}^T \Xi \left\{ \bar{\theta} - \Xi^{-1}\Upsilon \right\} \right.$$
$$\left. + \frac{1}{2N_\tau} Y_{0/\tau}^T \left\{ I - \Psi_{\tau/0}\Xi^{-1}\Psi_{0/\tau}^T \right\} Y_{\tau/0} \right)^2, \quad (25)$$

where $\Upsilon \in \mathbb{R}^n$ are defined by

$$\Upsilon \triangleq \Psi_{0/\tau}^T Y_{\tau/0} = \Psi_{0/\tau}^T Y_{\tau/0}. \quad (26)$$

Since the second term of (25) is independent of $\bar{\theta}$, (20) is unique and sufficient to minimize $J^2(\bar{\theta}, \tau, N)$ provided that A4 is satisfied. The minimum (21) of the cost function then follows directly by substituting (20) into (25). $\quad\square$

Equation (23) plays an important role which is similar to the normal equation of the method of least squares. Setting $\tau = 0$ and deleting the redundancy in (23) reduces to the normal equation of least-squares estimate.

Next, let the least-correlation estimate (20) be stated as

$$\hat{\theta}(\tau, N) = \left(\Psi_0^T \Psi_\tau + \Psi_\tau^T \Psi_0 \right)^{-1} \left(\Psi_0^T Y_\tau + \Psi_\tau^T Y_0 \right). \quad (27)$$

For $\tau = 0$, (27) specializes to the least-squares estimate

$$\hat{\theta}(0, N) = \left(\Psi^T \Psi \right)^{-1} \Psi^T Y, \quad (28)$$

where $\Psi \triangleq \Psi_0(0, N) = \Psi_\tau(0, N)$ and $Y \triangleq Y_0(0, N) = Y_\tau(0, N)$. Deleting the first term in each parenthesis of (27) yields the instrumental variable (IV) estimate

$$\hat{\theta}_{IV}(\tau, N) = \left(\Psi_\tau^T \Psi_0 \right)^{-1} \Psi_\tau^T Y_0 \quad (29)$$

with the instrumental variable Ψ_τ. This kind of instrumental variable estimates is useful for EIV problems [14, 15, 17] and that the 'bias eliminated least-squares (BELS)' approaches [21, 22] to EIV models can be interpreted as a sort of instrumental variable estimates [5, 6, 16].

One the other hand, deleting the second terms instead of the first terms in each parenthesis of (27) yields

$$\hat{\theta}_{IN1}(\tau, N) = \left(\Psi_0^T \Psi_\tau \right)^{-1} \Psi_0^T Y_\tau. \quad (30)$$

The other combinations of the terms in each parenthesis of (27) gives following induced estimators.

$$\hat{\theta}_{IN2}(\tau, N) = \left(\Psi_0^T \Psi_\tau \right)^{-1} \Psi_\tau^T Y_0, \quad (31)$$

$$\hat{\theta}_{IN3}(\tau, N) = \left(\Psi_\tau^T \Psi_0 \right)^{-1} \Psi_0^T Y_\tau. \quad (32)$$

These induced estimates (30)-(32) are slightly different from the instrumental variable method [8, 9].

IV. CONSISTENCY OF THE LCE

The parameter estimation error

$$\tilde{\theta}(\tau, N) \triangleq \hat{\theta}(\tau, N) - \theta \quad (33)$$

is stated as

$$\tilde{\theta}(\tau, N) = \left\{ \bar{R}_{\psi\psi}(t, t - \tau, N) + \bar{R}_{\psi\psi}(t - \tau, t, N) \right\}^{-1}$$
$$\times \left\{ \bar{r}_{\psi e}(t, t - \tau, N) + \bar{r}_{\psi e}(t - \tau, t, N) \right\}, \quad (34)$$

where $\bar{r}_{\psi e}(t_1, t_2, N) \in \mathbb{R}^n$ is defined by

$$\bar{r}_{\psi e}(t_1, t_2, N) \triangleq \frac{1}{N - \tau} \sum_{t=1+\tau}^N \psi(t_1)e(t_2) \quad (35)$$

with either $t_1 = t, t_2 = t - \tau$ or $t_1 = t - \tau, t_2 = t$ and $\tau \triangleq |t_1 - t_2|$. Correlation (35) is evaluated as

$$\bar{r}_{\psi e}(t_1, t_2, N) = \bar{r}_{\psi\eta}(t_1, t_2, N) - \bar{R}_{\psi\zeta}(t_1, t_2, N)\theta, \quad (36)$$

where $\bar{R}_{\psi\zeta}(t_1, t_2, N)$ and $\bar{r}_{\psi\eta}(t_1, t_2, N)$ are defined similarly to (10) and (35), respectively. Observations on (34) gives following property.

Theorem 3. (*Consistency*) Suppose that A1-A4 are satisfied. Then $\hat{\theta}(\tau, N)$ converges to θ with probability 1 as N increases toward infinity, that is,

$$\lim_{N\to\infty} \hat{\theta}(\tau, N) = \theta. \quad (37)$$

Proof. According to the ergodic theory [9, Theorem 2.3 in p.43], the empirical correlation $\bar{R}_{\psi\psi}(t_1, t_2, N)$ converges to the corresponding mathematical correlation $R_{\psi\psi}(t_1 - t_2)$ with probability 1 as N goes to infinity, that is,

$$\lim_{N\to\infty} \frac{1}{N - \tau} \sum_{t=1+\tau}^N \psi(t_1)\psi^T(t_2) = \bar{E}\left[\psi(t_1)\psi^T(t_2) \right] \quad (38)$$

or equivalently

$$\lim_{N\to\infty} \bar{R}_{\psi\psi}(t_1, t_2, N) = R_{\psi\psi}(t_1 - t_2). \quad (39)$$

Similarly as N increases to infinity, (36) comes to

$$r_{\psi e}(t_1 - t_2) = r_{\psi\eta}(t_1 - t_2) - R_{\psi\zeta}(t_1 - t_2)\theta \quad (40)$$

and then (34) is written as

$$\lim_{N\to\infty} \tilde{\theta}(\tau, N) = R_{\psi\psi}^{-1}(\tau) \left\{ r_{\psi\eta}(\tau) - R_{\psi\zeta}(\tau)\theta \right\} \quad (41)$$

since the mathematical correlations depend on the absolute difference of time, $\tau = |t_1 - t_2|$, owing to A2. Finally A3 guarantees that both $r_{\psi\eta}(\tau)$ and $R_{\psi\zeta}(\tau)$ are zero, which proves (37). $\quad\square$

Corollary 4. In the EIV model (1)-(5), the least-squares estimates $\hat{\theta}(0, N)$ in (28) yields the error-prone results given by

$$\lim_{N\to\infty} \tilde{\theta}(0, N) = R_{\psi\psi}^{-1}(0) \left\{ r_{\psi\eta}(0) - R_{\psi\zeta}(0)\theta \right\}. \quad (42)$$

Proof. As $\hat{\theta}(\tau, N)$ comes to $\hat{\theta}(0, N)$ at $\tau = 0$, letting $\tau = 0$ for (41) yields (42). $\quad\square$

The cross-correlation matrix $R_{\psi\zeta}(0)$ in (42) does not degenerate to zero any more even if $\zeta(t)$ is white. Therefore the least-square estimate can never be consistent for EIV models.

Corollary 5. Suppose that A1-A3 are satisfied and either $\bar{R}_{\psi\psi}(t, t - \tau, N)$ or $\bar{R}_{\psi\psi}(t - \tau, t, N)$ corresponding to each of (29)-(32) has a full rank. Then each of (29)-(32) converges to the true parameter with probability 1 as N increases toward infinity.

Proof. As N goes to infinity, each of the induced estimates converges to the same value with (41), which is deduced from the proof of Theorem 3. $\quad\square$

V. Geometrical Aspect of the LCE

Consider (23). Following two equations

$$\Psi_{0/\tau}^T Y_{\tau/0} - \Psi_{0/\tau}^T \Psi_{\tau/0}\hat{\theta}(\tau) = 0, \qquad (43)$$

$$\Psi_{\tau/0}^T Y_{0/\tau} - \Psi_{\tau/0}^T \Psi_{0/\tau}\hat{\theta}(\tau) = 0, \qquad (44)$$

are equivalent to each other. Let $\hat{Y}_{\tau/0}$ and $\hat{Y}_{0/\tau}$ be

$$\hat{Y}_{\tau/0}\left(\hat{\theta}(\tau)\right) = \Psi_{\tau/0}\hat{\theta}(\tau), \qquad (45)$$

$$\hat{Y}_{0/\tau}\left(\hat{\theta}(\tau)\right) = \Psi_{0/\tau}\hat{\theta}(\tau) \qquad (46)$$

and let the corresponding residuals be

$$\mathcal{E}_{\tau/0}\left(\hat{\theta}(\tau)\right) = Y_{\tau/0} - \Psi_{\tau/0}\hat{\theta}(\tau), \qquad (47)$$

$$\mathcal{E}_{0/\tau}\left(\hat{\theta}(\tau)\right) = Y_{0/\tau} - \Psi_{0/\tau}\hat{\theta}(\tau). \qquad (48)$$

From above expressions we can get the following property called the principle of orthogonality.

Lemma 6. Suppose that A4 is satisfied. $J^2(\bar{\theta}, \tau)$ achieves minimum if and only if

$$\Psi_{0/\tau}^T \mathcal{E}_{\tau/0} = \Psi_{\tau/0}^T \mathcal{E}_{0/\tau} = 0. \qquad (49)$$

Proof. Each of (43) and (44) is necessary and sufficient for $\hat{\theta}(\tau)$ to minimize $J^2(\bar{\theta}, \tau)$ provided that A4 is satisfied. Using (47)-(48) to (43)-(44) yields (49). □

Corollary 7. The equation (49) is equivalent to

$$\hat{Y}_{0/\tau}^T \mathcal{E}_{\tau/0} = \hat{Y}_{\tau/0}^T \mathcal{E}_{0/\tau} = 0. \qquad (50)$$

Proof. Left-multiplying $\hat{\theta}^T(\tau)$ to (49) and employing (45)-(46) yields (50). □

Suppose that there exist linear vector spaces $\mathbf{Y}_{\tau/0}$ and $\mathbf{Y}_{0/\tau}$ spanned by $Y_{\tau/0}$ and $Y_{0/\tau}$, respectively, and their corresponding subspaces $\hat{\mathbf{Y}}_{\tau/0}$ and $\hat{\mathbf{Y}}_{0/\tau}$ spanned by $\hat{Y}_{\tau/0}$ and $\hat{Y}_{0/\tau}$. Let $P_{\tau/0}$ and $P_{0/\tau}$, defined by

$$P_{\tau/0} \triangleq \Psi_{0/\tau}\left\{\Psi_{\tau/0}^T\Psi_{0/\tau}\right\}^{-1}\Psi_{\tau/0}^T, \qquad (51)$$

$$P_{0/\tau} \triangleq \Psi_{\tau/0}\left\{\Psi_{0/\tau}^T\Psi_{\tau/0}\right\}^{-1}\Psi_{0/\tau}^T, \qquad (52)$$

operate on $\mathbf{Y}_{\tau/0}$ and $\mathbf{Y}_{0/\tau}$, respectively. Then $P_{\tau/0}$ and $P_{0/\tau}$ map the measured responses, $Y_{\tau/0}$ and $Y_{0/\tau}$, to their estimates, $\hat{Y}_{\tau/0}$ and $\hat{Y}_{0/\tau}$, respectively. That is,

$$\hat{Y}_{\tau/0}\left(\hat{\theta}(\tau)\right) = P_{\tau/0}Y_{\tau/0}, \qquad (53)$$

$$\hat{Y}_{0/\tau}\left(\hat{\theta}(\tau)\right) = P_{0/\tau}Y_{0/\tau} \qquad (54)$$

which are obtained by using (20) into (45) and (46). For discussion in Lemma 8 we introduce

$$J = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}, \qquad (55)$$

where $I \in \mathbb{R}^{(N-\tau)\times(N-\tau)}$ is the identity matrix.

Lemma 8. $P_{\tau/0}$ and $P_{0/\tau}$ have following properties:
1. Both $P_{\tau/0}$ and $P_{0/\tau}$ are oblique projection operators.
2. $P_{\tau/0} = P_{0/\tau}^T$ and $P_{0/\tau} = P_{\tau/0}^T$.
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 $(I - P_{\tau/0})J = J(I - P_{0/\tau})$.

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$[J(I - P_{0/\tau})]$.

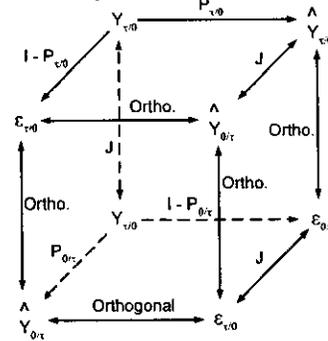The properties and the relationships stated in Lemma 8 can be depicted in Fig. 2.

Fig. 2. Relationship among spaces and corresponding operators

Proof. Each item is composed of two similar or symmetric states. We will sketch only one of each.
1. Each operator is idempotent, but not self-adjoint [2] [4, p.71], i.e., $P_{\tau/0}^2 = P_{\tau/0}$, but $P_{\tau/0} \neq P_{\tau/0}^T$.
2. It is self-evident from (51) and (52).
3. Applying the equivalent expression

$$P_{0/\tau} = \begin{bmatrix} \Psi_\tau\Xi^{-1}\Psi_0^T & \Psi_\tau\Xi^{-1}\Psi_\tau^T \\ \Psi_0\Xi^{-1}\Psi_0^T & \Psi_0\Xi^{-1}\Psi_\tau^T \end{bmatrix} \qquad (56)$$

to the right-hand side of the first equality shows

$$JP_{0/\tau}J = \Psi_{0/\tau}\Xi^{-1}\Psi_{\tau/0}^T = P_{\tau/0}. \qquad (57)$$

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

$$\left[(I - P_{0/\tau})J\right]^T = J - JP_{\tau/0}JJ = (I - P_{0/\tau})J, \qquad (58)$$

where we use the fact $I = JJ$.
6. Consider an arbitrary $Y_{\tau/0} \in \mathbf{Y}_{\tau/0}$. Applying each operator $P_{\tau/0}$ and $(I - P_{0/\tau})J$ to $Y_{\tau/0}$ generates $\hat{Y}_{\tau/0}$ and

$$(I - P_{0/\tau})JY_{\tau/0} = Y_{0/\tau} - \hat{Y}_{0/\tau} = \mathcal{E}_{0/\tau}, \qquad (59)$$

respectively. According to Corollary 7, $\mathcal{E}_{0/\tau}$ and $\hat{Y}_{\tau/0}$ are orthogonal. Thus the spaces spanned by $\mathcal{E}_{0/\tau}$ and $\hat{Y}_{\tau/0}$ are also orthogonal. □

VI. Numerical Examples

It is known that the Fourier series with finite elements is an optimal fitting in the sense of least squares [1]. We modify the problem into the EIV setting with

$$\phi(t) = \begin{bmatrix} \sin 2\pi t & \sin 6\pi t \end{bmatrix}^T$$

$$\theta = \begin{bmatrix} 4/\pi & 4/3\pi \end{bmatrix}^T$$

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

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

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

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

2920

where $\xi_\eta(t)$ and $\xi_{\zeta_i}(t)$ are white and the variance of each of them is chosen such that the signal-to-noise ratios

$$\mathrm{SNR}_i = 10\log_{10}\left(\bar{E}\left[\phi^T(t)\phi(t)\right]/E\left[\zeta^T(t)\zeta(t)\right]\right) \quad (60)$$

$$\mathrm{SNR}_o = 10\log_{10}\left(\bar{E}\left[z^2(t)\right]/E\left[\eta^2(t)\right]\right) \quad (61)$$

are about 20dB, 10dB, 5dB, or 1dB, but the maximum SNR_o is limited to about 10dB due to $\eta_1(t)$. According to A3, $\zeta(t)$ 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:

$$\text{Case 1. FIR} \quad : \quad G_\zeta(q^{-1}) = 0.3 + 0.7q^{-1} \quad (62)$$

$$\text{Case 2. IIR} \quad : \quad G_\zeta(q^{-1}) = \frac{0.3}{1 - 0.7q^{-1}}. \quad (63)$$

TABLE I
EFFECT OF SNR - FIR NOISE, 10000 SAMPLES

Estimate	SNR_i	SNR_o	$\check{\theta}_1$ [%]	$\check{\theta}_2$ [%]
$\hat{\theta}(0)$	20dB	10dB	-1(-1.0) ± 0	-1(-1.3) ± 0
	10	10	-9(-9.0) ± 1	-9(-9.3) ± 0
	5	5	-23(-24.0) ± 1	-23(-24.2) ± 4
	1	1	-45(-44.1) ± 2	-45(-44.3) ± 6
$\hat{\theta}(\tau)$	20	10	0 ± 0	0 ± 0
$(\tau = 2)$	10	10	0 ± 1	0 ± 1
	5	5	1 ± 1	1 ± 5
	1	1	-2 ± 2	-2 ± 11

TABLE II
EFFECT OF N - FIR NOISE, $\mathrm{SNR}_i \simeq \mathrm{SNR}_o \simeq 5\mathrm{dB}$

Estimate	# of Data	$\check{\theta}_1$ [%]	$\check{\theta}_2$ [%]
$\hat{\theta}(0)$	100	-25(-24.2) ± 14	-15(-24.4) ± 37
	1000	-21(-24.1) ± 6	-26(-24.3) ± 10
	10000	-23(-24.0) ± 1	-23(-24.2) ± 4
	100000	-24(-24.0) ± 0	-24(-24.2) ± 1
$\hat{\theta}(\tau)$	100	-1 ± 19	16 ± 42
$(\tau = 2)$	1000	4 ± 8	-6 ± 17
	10000	1 ± 1	1 ± 5
	100000	0 ± 1	-1 ± 1

TABLE III
INDUCED ESTIMATORS - FIR NOISE, $\mathrm{SNR}_i \simeq \mathrm{SNR}_o \simeq 5\mathrm{dB}$. 10000
SAMPLES

Estimate ($\tau = 2$)	$\check{\theta}_1$ [%]	$\check{\theta}_2$ [%]
$\hat{\theta}(\tau)$	1 ± 1	1 ± 5
$\hat{\theta}_{IV}(\tau)$	1 ± 1	2 ± 4
$\hat{\theta}_{IN1}(\tau)$	1 ± 1	0 ± 6
$\hat{\theta}_{IN2}(\tau)$	1 ± 1	-1 ± 6
$\hat{\theta}_{IN3}(\tau)$	1 ± 1	4 ± 4

Table I-IV summarizes the results from 100 Monte Carlo runs for each case. Each table shows the estimation errors defined by

$$\check{\theta}_i = \frac{1}{|\theta_i|}\left\{\bar{\bar{\tilde{\theta}}}_i \pm \bar{\sigma}(\tilde{\theta}_i)\right\}, \quad i = 1, 2, \quad (64)$$

where $\bar{\bar{\tilde{\theta}}}_i$ and $\bar{\sigma}(\tilde{\theta}_i)$ denote the empirical mean and the empirical standard deviation of $\tilde{\theta}_i$, respectively. The values in parentheses of each table are evaluated by (41) or (42). For the IIR noise in Table IV, the correlations with sufficiently

large time intervals are counted in. Generally speaking there are good agreements between the theoretical calculations and the Monte Carlo simulations.

TABLE IV
EFFECT OF τ ON IIR NOISE - $\mathrm{SNR}_i \simeq \mathrm{SNR}_o \simeq 5\mathrm{dB}$, 10000 SAMPLES

Estimate	τ	$\check{\theta}_1$ [%]	$\check{\theta}_2$ [%]
$\hat{\theta}(0)$	-	-24(24.0) ± 1	-24(24.2) ± 4
$\hat{\theta}(\tau)$	2	-14(-13.5) ± 2	-14(-14.3) ± 5
	4	-7(-7.3) ± 2	-10(-9.4) ± 5
	8	-2(-2.0) ± 2	-23(-22.3) ± 22
	16	0(-0.2) ± 2	-1(0.1) ± 6
	32	0(0.0) ± 3	-2(0.0) ± 5
	64	0(0.0) ± 2	0(0.0) ± 6

Table I and Table II confirm numerically that the method of least correlation works well for the EIV problem. Table II says that the LCE has a capability to find out true parameters from severely contaminated data provided that the number of data samples is sufficiently large. Table III summarizing the performance of four induced estimators shows that all of them generate reasonable results. It is expected that the instrumental variable estimator $\hat{\theta}_{IV}(\tau)$ works on the EIV model as known in literature [14, 15, 17], but it is interesting for the others $\hat{\theta}_{IN_i}(\tau), i = 1, 2, 3$ to make comparable estimates with $\hat{\theta}_{IV}(\tau)$. We show in Table IV that the method of least correlation can be applied to the problems with regressors corrupted by infinitely correlated noise which apparently violates A3. Table IV says that the least-correlation method can give pretty good estimates if A4 is satisfied with a sufficiently large τ. The simulation results in Table IV are also supported by the theoretical calculations.

VII. RECURSIVE LEAST-CORRELATION ALGORITHMS

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

$$\psi_{i/i-\tau} \triangleq \begin{bmatrix} \psi(i) & \psi(i-\tau) \end{bmatrix}$$

$$\psi_{i-\tau/i} \triangleq \begin{bmatrix} \psi(i-\tau) & \psi(i) \end{bmatrix}$$

$$y_{i-\tau/i} \triangleq \begin{bmatrix} y(i-\tau) & y(i) \end{bmatrix}^T$$

respectively, an equivalent expression of (20) is written as

$$\hat{\theta}(\tau, t) = \left(\sum_{i=1+\tau}^{t} \psi_{i/i-\tau}\psi_{i-\tau/i}^T\right)^{-1}\sum_{i=1+\tau}^{t} \psi_{i/i-\tau}y_{i-\tau/i}. \quad (65)$$

Employing the steps [8, pp.262-263], which derives the RLS (recursive least-squares) algorithm from its off-line version - the least-squares estimate (28), for the least-correlation estimate (20) yields the RLC (recursive least-correlation) algorithm

$$\hat{\theta}(\tau, t) = \hat{\theta}(\tau, t - 1)$$
$$+K(t)\left(y_{t-\tau/t} - \psi_{t-\tau/t}^T\hat{\theta}(\tau, t - 1)\right) \quad (66)$$

$$K(t) = P_{t-1}\psi_{t/t-\tau}\left(I + \psi_{t-\tau/t}^T P_{t-1}\psi_{t/t-\tau}\right)^{-1} \quad (67)$$

$$P_t = P_{t-1} - K(t)\psi_{t-\tau/t}^T P_{t-1} \quad (68)$$

2921

for $t > \tau$ provided that P_τ and $\hat{\theta}(\tau, \tau)$ are given. The recursive algorithm (66)-(68) is equivalent to (20), which means that the estimates from both of them are same at the final time. Fig. 3 shows the estimate for the Example 1 in Section VI by the RLC algorithm (66)-(68) and the RLS algorithm [8,9].
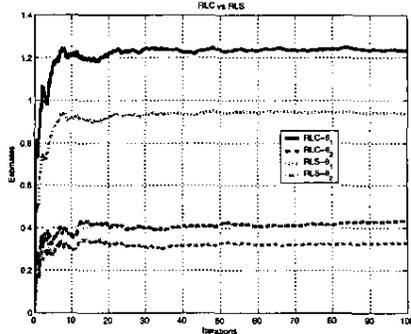

Fig. 3. Online estimates by RLS and RLC algorithms : Example 1 ($\theta_1 = 4/\pi, \theta_2 = 4/3\pi$) - FIR input noise, $\tau = 2$, $SNR_i = SNR_o = 5dB$

On the other hand, for the continuous-time version of (14)

$$J^2(\bar{\theta}, \tau, t) = \left(\frac{1}{t-\tau} \int_\tau^t \epsilon(s, \bar{\theta})\epsilon(s - \tau, \bar{\theta})ds \right)^2, \quad (69)$$

applying the minimization procedure [13, pp. 370-371] gives the algorithm - continuous-time RLC algorithm

$$\dot{P}(t) = -P(t)\left[\psi(t)\psi^T(t-\tau) + \psi(t-\tau)\psi^T(t)\right]P(t) \quad (70)$$

$$\dot{\theta}(t) = -P(t)\left[\psi(t)e(t-\tau|t) + \psi(t-\tau)e(t|t)\right] \quad (71)$$

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

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

Note that t in (69)-(71) denotes the time on continuous-time domain.

VIII. CONCLUDING REMARKS

Based on observations about the residuals which are resulted from the least-squares estimate applied to EIV models, we introduce a criterion defined by the square of empirical correlation between residuals. The necessary and sufficient condition minimizing the criterion yields the least-correlation estimate. Analysis concludes that the estimate converges to the true value as the number of samples increases toward infinity. Monte Carlo simulations support the analysis. Moreover, the numerical results hint the capability of the estimate to deal with the infinitely correlated noise. The least-correlation-based estimator has some interesting geometrical properties which are partly similar to the least-squares estimate and partly different from. Finally we mention briefly the recursive least-correlation algorithms on continuous-time domain as well as on discrete-time domain.

It is possible to interpret the least-correlation estimate as a deterministic representation of the 'error whitening Wiener filter' [10]. The previous works [10, 11] states the stochastic expressions of the unbiasedness and the orthogonality. In this paper, we discuss many other aspects of

the least-correlation estimate, for axample, the sufficiency for minimum and the induced estimators in Section III, the consistency in Section IV, the geometrical interpretations in Section V, and the recursive realization on continuous-time domain in Section VII.

This work can be extended to the estimates for EIV nonlinear models [7]. Input design problem, which is closely related to A4, for the best estimates is a further work for application. Another further study will be how to realize the RLS algorithm numerically robust and computationally efficient.

REFERENCES

[1] J. S. Bay, *Fundamentals of Limear Space Systems*, McGraw-Hill, 1999.
[2] R. T. Behrens and L. L. Scharf, "Signal Processing Applications of Oblique Projection Operators," *IEEE Trans. Signal Processing*, Vol. 42, No. 6, pp. 1413-1424, June 1994.
[3] C. B. Feng and W, X. Zheng, "Robust Identification of Stochastic Linear Systems with Correlated Noise," *IEE Proceedings, Part D*, Vol. 138, No. 5, pp. 484-492, Sep. 1991.
[4] P. A. Fuhrmann, *Linear Systems and Operators in Hilbert Space*, McGraw-Hill, 1981.
[5] L.-J. Jia, T. Hanada, C.-Z. Jim, Z.-J. Yang, and K. Wada, "On the Relationship between BELS and IV Methods," *Proceedings of the 41st IEEE Conference on Decision and Control*, pp. 3440-3445, Dec. 2002.
[6] L.-J. Jia, T. Hanada, C.-Z. Jim, Z.-J. Yang, and K. Wada, "A Unified Form for BELS Method in Colored Noise Case," *Proceedings of the 41st IEEE Conference on Decision and Control*, pp. 2637-2644, Dec. 2002.
[7] B. E. Jun and D. S. Bernstein, "Least-Correlations Estimates for Errors-in-Variables Nonlinear Models," IECON 2004.
[8] R. Johansson, *System Modeling and Identification*, Prentice-Hall, New Jersey, 1993.
[9] L. Ljung, *System Identification - Theory for the User*, 2nd Ed., Prentice-Hall, New Jersey, 1999.
[10] J. C. Principe, Y. N. Rao and D. Erdogmus, "Error Whitening Wiener Filters: Theory and Algorithms," In S. Kaykin and B. Widrow (Eds.) *Advances in LMS Filters, Chap. 10*, New York, NY: Wiley, 2003.
[11] Y. N. Rao, D. Erdogmus, G. Y. Rao and J. C. Principe, "Stochastic Error Whitening Algorithm for Linear Filter Estimation with Noisy Data," *Neural Networks*, Vol. 16, pp. 873-990, 2003.
[12] W. Scherrer and M. Deistler, "A Structure Theory for Linear Dynamic Errors-in-Variables Models," *SIAM J. Control Optimization*, Vol. 36, No. 6, pp. 2148-2175, Nov. 1998.
[13] J. E. Slotine and W. Li, *Applied Nonlinear Control*, Prentice-Hall, New Jersey, 1991.
[14] T. Soderstrom, U. Soverini, and K. Mahata, "Perspectives on Errors-in-variables Estimation for Dynamic Systems," *Signal Processing*, Vol. 82, pp. 1139-1154, 2002.
[15] T. Soderstrom and K. Mahata, "On Instrumental Variable and Total Least Squares Approaches for Identification of Noisy Systems," *Int. J. Control*, Vol. 75, No. 6, pp. 381-389, 2002.
[16] T. Soderstrom, W. X. Zheng, and P. Stoica, "Comments on "On a Least-Squares-Based Algorithm for Identification of Stochastic Linear Systems"," *IEEE Trans. Signal Processing*, Vol. 47, No. 5, pp. 1395-1396, May 1999.
[17] P. Stoica, M. Cedervall, and A. Eriksson, "Combined Instrumental Variable and Subspace Fitting Approach to Parameter Estimation of Noisy Input-Output Systems," *IEEE Trans. Signal Processing*, Vol. 43, No. 10, pp. 2386-2397, Oct. 1995.
[18] S. van Huffel and P. Lemmerling (Ed.), *Total Least Squares Techniquea and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, Kluwer Academic Publisher, 2002.
[19] S. van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, 1991.
[20] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*, AIAA, 2000.
[21] W. X. Zheng, "A Bias Correction Method for Identification of Linear Dynamic Errors-in-Variables Methods," *IEEE Trans. Automatic Control*, Vol. 47, No. 7, pp. 1142-1147, July 2002.
[22] W. X. Zheng, "On a Least-Squares-Based Algorithm for Identification of Stochastic Linear Systems," *IEEE Trans. Signal Processing*, Vol. 46, pp. 1631-1638, 1998.