PREDICTION ERROR VS SUBSPACE METHODS IN CLOSED LOOP IDENTIFICATION

Alessandro Chiuso [*,1] **Giorgio Picci** [*,1]

[*] *Department of Information Engineering, University of Padova, 35131 Padova, Italy*
e-mail:{*chiuso,picci*}*@dei.unipd.it*

Abstract: In this paper we investigate the role of the output predictor in subspace identification. We shall see that subspace identification of the predictor model can ideally yield consistent estimates regardless of the presence of feedback. This solves a longstanding open question in system identification.

Keywords: Subspace identification, closed loop identification, prediction error methods.

1. REMARKS ON CLOSED-LOOP PREDICTION ERROR IDENTIFICATION

Let $\mathbf{z} := [\mathbf{u}\ \ \mathbf{y}]^\top$ be a wide sense stationary zero mean process, which is second-order ergodic and with a full-rank rational spectral density matrix. There may be feedback from $\mathbf{y}$ to $\mathbf{u}$, see (Granger 1963, Caines and Chan 1976, Gevers and Anderson 1981) for a definition of this concept.

We discuss the identification of the "deterministic" transfer function $F(z)$ of the linear stationary (innovation) model of $\mathbf{y}$ given $\mathbf{u}$, of the usual form

$$\mathbf{y}(t) = F(z)\mathbf{u}(t) + G(z)\mathbf{e}(t) \qquad (1.1)$$

where $F(z)$ and $G(z)$ are rational transfer matrices and $\mathbf{e}$ is the one step ahead prediction error $\mathbf{e}(t) = \mathbf{y}(t) - E\left[\mathbf{y}(t) \mid \mathcal{Z}_t^-\right]$, $\mathcal{Z}_t^-$ being the joint infinite past of $\mathbf{y}$ and $\mathbf{u}$ up to time t. Without loss of generality, we can assume that $F(\infty) = 0$. It is well-known that the input-output relation(1.1) is the innovation representation of $\mathbf{y}$ given $\mathbf{u}$, if and only if $G(z)$ is minimum phase (no zeros outside of the unit circle) and such that $G^{-1}(z)F(z)$ has no unstable poles (Ljung 1997).

Early results in the literature, (Caines and Chan 1976, Gevers and Anderson 1981, Ng *et al.* 1977), which we shall here give for granted discuss identifiability of linear feedback models of this kind.

Two main difficulties in closed-loop identification are the correlation of the white noise $\mathbf{e}$ with (past) inputs $\mathbf{u}$ and the possible instability of the open-loop plant $F(z)$ to be identified. Both difficulties are circumvented in PEM identification by *identifying the predictor* $\hat{\mathbf{y}}_\theta(t|t-1) := E_\theta[\mathbf{y}(t) \mid \mathcal{Z}_t^-]$, which should be thought just as a deterministic (parameterized) linear dynamical system processing the past input data $\mathbf{y}$ and $\mathbf{u}$. The correlation problem has no impact in the parameter estimation phase as it influences only the structure of the predictor and can be dealt with theoretically beforehand. The instability also has virtually no influence since under mild assumptions on the data generating system (the true $F(z)$ may well be unstable!), *the predictor is an asymptotically stable system* (compare e.g. the asymptotic stability of the Kalman Filter) and the usual statistical

[1] Work supported by the Italian National Ministry of Higher Education (MIUR) and by the European Community project RECSYS.

asymptotics based on stationarity and ergodicity naturally applies.

Since there is basically a one to one correspondence between the predictor model and the pair $(F(z), G(z))$ all the above applies directly also to the identification of the model (1.1). It is in fact well-known and it has been recently been restated in (Forsell and Ljung 1999) that prediction error methods, provided the model class is (in a very specific sense) "rich enough", yield consistent estimators of $F(z)$ regardless of the presence of feedback.

On the contrary, until very recently subspace methods have not been able to cope successfully with the presence of feedback. It is the purpose of this paper to clarify this situation and to provide a possible explanation of this fact. We shall argue that subspace system identification can also naturally be restated as *predictor identification* and in this framework subspace methods could in principle yield the same kind of asymptotic behavior of PEM methods also with closed-loop data. As we shall see however, this is true only up to a point since, unlike PEM methods, subspace methods are sensitive to initial conditions and the necessity of working in practice with regressions on a finite amount of data prevents subspace procedures to yield consistent estimates.

1.1 Notations and background

The notations are fairly standard: boldface letters will denote (in general vector-valued) random quantities, which we invariably assume zero-mean and with a finite covariance matrix. Calligraphic capital symbols (possibly with time subscripts) denote linear subspaces of random variables endowed with the usual inner product $\langle \xi, \eta \rangle := E\{\xi\eta\}$, the operator E denoting mathematical expectation. The symbol $E[\zeta \mid \mathcal{X}]$ denotes the vector of orthogonal projections (conditional expectations in the Gaussian case) of the components of $\zeta \in \mathcal{Z}$ onto the subspace $\mathcal{X}$.

Let the subspaces $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{Z}$ be in direct sum, i.e. $\mathcal{A} \cap \mathcal{B} = \{0\}$, then the orthogonal projection of any element $\zeta \in \mathcal{Z}$ onto the direct sum $\mathcal{A} + \mathcal{B}$ can be written uniquely as a sum of elements of $\mathcal{A}$ and $\mathcal{B}$, namely

$$E\{\zeta \mid \mathcal{A} + \mathcal{B}\} = E_{||\mathcal{A}}\{\zeta \mid \mathcal{B}\} + E_{||\mathcal{B}}\{\zeta \mid \mathcal{A}\}$$

where $E_{||\mathcal{A}}\{\zeta \mid \mathcal{B}\}$ is called the *oblique projection of ζ onto $\mathcal{B}$ along $\mathcal{A}$* and $E_{||\mathcal{B}}\{\zeta \mid \mathcal{A}\}$ is called the *oblique projection of ζ onto $\mathcal{A}$ along $\mathcal{B}$*. The notation $\mathcal{A} \perp \mathcal{B} \mid \mathcal{C}$ means that the two subspaces $\mathcal{A}$ and $\mathcal{B}$ are *conditionally orthogonal given a third subspace* $\mathcal{C}$, namely, for any $\alpha \in \mathcal{A}$ and $\beta \in \mathcal{B}$

$$\langle \alpha - E\{\alpha \mid \mathcal{C}\}, \beta - E\{\beta \mid \mathcal{C}\} \rangle = 0$$

If $\mathcal{C} = \{0\}$, conditional orthogonality reduces to the usual orthogonality $\mathcal{A} \perp \mathcal{B}$.

In this paper, rather than working with notationally cumbersome finite arrays of observed sample data (e.g. finite Hankel matrices) and then taking limits as the sample length N tends to infinity, as it is often done in the subspace identification literature, we shall work entirely in a stochastic setting. Under the assumed second order ergodicity of the observed processes, for $N \to \infty$ the the sample covariances can be substituted by the true ones and we can essentially rephrase identification in terms of random variables rather than numerical data. For a more precise description of the equivalence of the two setups see (Lindquist and Picci 1996).

Still, in order to deal with realistic subspace algorithms which can only regress on a finite amount of data, we shall keep *finite past and future horizons* (the "i" parameter of (Van Overschee and De Moor 1994) or the p and f parameters in most subsequent subspace literature). This setting we shall describe as using data from a *finite observation interval* later on. In this paper finite (and generally fixed) past and future horizons will hold even when the sample size N is let going to ∞ for the purpose of asymptotic analysis. Because of this intrinsic limitation, the effect of initial conditions has to be taken into account and will, as we shall see later, lead to generally biased estimates. This unpleasant effect of finite observation interval on the estimates does not show when there is no feedback.

Even if the effect of initial conditions (and hence the bias) could theoretically be eliminated by letting the past horizon $p := t - t_0$ tend to infinity (perhaps at at a certain rate), we shall remind the reader that in practice the regression estimates in subspace identification involve sample covariance matrices of past and future horizon data, estimated with finite data length N. It is well-known that the variance of these sample covariances grows (for fixed N) rather sensibly when the past and future horizons p and f are expanded and it is a good general rule of thumb not to increase p and f over a certain threshold, which depends on N.

2. SUBSPACE IDENTIFICATION VIA PREDICTOR MODELS

Naturally, subspace methods are designed for the identification of state space models of the form

$$\begin{cases} \mathbf{x}(t+1) &= A\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{e}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{e}(t) \end{cases} \tag{2.1}$$

which we assume is a minimal state space realization of the I/O model (1.1), so that $F(z) = C(zI -$

$A)^{-1}B$, $G(z) = C(zI - A)^{-1}K + I$. It is shown in the stochastic realization literature, see (Chiuso and Picci 2003), that the *state space* of (2.1) (the subspace spanned by the scalar components of the state vector $\mathbf{x}(t)$), can abstractly be constructed as the space spanned by the oblique projections of the future outputs $\{\mathbf{y}(t+k),\, k = 0, 1, \dots\}$ onto the "joint past inputs" ($\mathbf{u}(s)$ and $\mathbf{e}(s)$, $s < t$) along the future "joint inputs" ($\mathbf{u}(s)$ and $\mathbf{e}(s)$, $s \geq t$) of the model (2.1), namely

$$\mathcal{X}_t^{+/-} = E_{\| \mathcal{U}_t^+ \vee \mathcal{E}_t^+} \left[\mathcal{Y}_t^+ \mid \mathcal{U}_t^- \vee \mathcal{E}_t^- \right]. \qquad (2.2)$$

This object is called an *oblique predictor space* (Picci 1997). As discussed in (Chiuso and Picci 2003, Chiuso and Picci 2005) this recipe for constructing the state space works in the presence of feedback , *provided the transfer function $F(z)$ is stable*. There are troubles with this construction when the "deterministic" transfer function $F(z)$ has unstable dynamics since in this case the joint past and future input spaces intersect and the oblique projection looses its meaning.

Now, it is apparent from the early literature, from (Van Overschee and De Moor 1993) to (Chiuso and Picci 2004*a*), that we can abstractly regard most subspace identification methods as different implementations of a basic two step procedure:

(1) Construct the state spaces $\mathcal{X}_t^{+/-}$ and $\mathcal{X}_{t+1}^{+/-}$ and choose bases $\mathbf{x}(t)$ and $\mathbf{x}(t+1)$,
(2) Compute the system matrices A, B, C, K from the chosen bases. In practice, given $\mathbf{x}(t)$ and $\mathbf{x}(t+1)$, this amounts to solving a linear regression for the unknown parameters in (2.1) by least squares [2].

Remark 2.1 The first step is essentially common to all subspace algorithms. In (Chiuso and Picci 2004*a*) the estimation of the observability matrix as done in the MOESP class of algorithms (Verhaegen 1994) has been shown to be equivalent to regressing on the state as in step (1) (the so-called "state approach"). The second step can sometimes be implemented differently but this does not change the essence of our discussion. Later we shall only refer to the "state approach". We just mention here that, under some specific circumstances, a "state-approach" based predictor identification algorithm can be shown to be asymptotically efficient (Chiuso 2005). ◇

It is really the specific procedure adopted to construct $\mathcal{X}_t^{+/-}$ which makes for the merits and demerits of each method. It is a fact that all standard subspace procedures (see (Larimore 1990, Van Overschee and De Moor 1994, Verhaegen 1994, Picci and Katayama 1996, Chiuso and Picci 2004*b*)) fail when data are collected in closed loop since the state construction step is based on some sort of manipulation of the model (2.1). In other words, these procedures attempt to construct the state space from the equations (2.1). This invariably requires that $\mathbf{e}(t)$ should be orthogonal to the whole input history $\mathcal{U}$ which is equivalent to absence of feedback from $\mathbf{y}$ to $\mathbf{u}$ (see (Granger 1963)). For more comments on these procedures see also the discussion in (Ljung and McKelvey 1996).

It should be said that (Chou and Verhaegen 1997, Van Overschee and De Moor 1997) formally deal with closed-loop systems. However the proposed algorithms either need some extra data (say the Markov parameters of the controller) or are extremely sensitive to noise. Other approaches, require some sort of preliminary ARX modeling, either directly (Ljung and McKelvey 1996) or in order to remove undesirable terms due to feedback (Jansson 2003). Also the algorithm of (Qin and Ljung 2003), which uses that fact that $\mathcal{X}^{+/-}$ can be obtained via the oblique projection (2.2), does not require orthogonality of innovations and input variables but turns out to be very sensitive to instability of $F(z)$.

What we would like to stress is that this state of affairs should not be seen as an intrinsic limitation of subspace methods in the presence of feedback, but rather should be attributed to the way the state space is constructed.

In our recent work (Chiuso and Picci 2005), inspired by an idea of (Jansson 2003), we have suggested an alternative procedure to construct the oblique predictor space $\mathcal{X}_t^{+/-}$ which does not require extra data, and does not suffer from the possible ill-conditioning mentioned above, occurring when the open loop system is unstable. It is based on the idea of looking at the *inverse system generating the innovations* $\mathbf{e}(t)$ from the joint process $\mathbf{y}$ and $\mathbf{u}$, namely the "whitening filter realization"

$$\begin{cases} \mathbf{x}(t+1) &= \bar{A}\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{y}(t) \\ \mathbf{e}(t) &= -C\mathbf{x}(t) + \mathbf{y}(t) \end{cases} \qquad (2.3)$$

where $\bar{A} := A - KC$. Here the state process $\mathbf{x}(t)$ is the same as in (2.1) so that *the two models have the same state space* $\mathcal{X}_t^{+/-}$ (a very well-known fact!). This model is well-known to be asymptotically stable under mild conditions on the zeros of the system.

Remark 2.2 Since $\mathcal{X}_t^{+/-}$ is the state space of the system producing the innovation $\mathbf{e}(t)$ from past input and output measurements $\{\mathbf{y}(s), \mathbf{u}(s), s \leq t\}$, it follows by the general recipe described before, that $\mathcal{X}_t^{+/-}$ must be the oblique predictor space of $\mathcal{E}_t^+$ given $\mathcal{Z}_t^+$, i.e.

[2] Actually to estimate the stochastic parameters of the model it also takes solving a Riccati equation.

$$\mathcal{X}_t^{+/-} = E_{\|\mathcal{Z}_t^+}\left[\mathcal{E}_t^+ \mid \mathcal{Z}_t^-\right]. \qquad (2.4)$$

This oblique projection of course makes sense if and only if $\mathcal{Z}_t^+ \cap \mathcal{Z}_t^- = \{0\}$, which is guaranteed if the spectrum of the joint process is bounded away from zero (Hannan and Poskitt 1988). For finite dimensional models this in particular requires that $\bar{A} := A - KC$ be strictly stable, i.e. there should be no zeros of the noise filter on the unit circle [3]. ◇

An observation made in (Chiuso and Picci 2005) is that one need not pre-compute the future innovation space $\mathcal{E}_t^+$ to obtain $\mathcal{X}_t^{+/-}$. The following result is quoted from (Chiuso and Picci 2005)

Theorem 2.1. Assume that the joint process satisfies

$$\mathcal{Z}_t^+ \cap \mathcal{Z}_t^- = \{0\}$$

then the space $\mathcal{X}_t^{+/-}$ is generated by the oblique projections $E_{\|\mathcal{Z}_{[t,t+k)}}\left[\mathcal{Y}_{t+k} \mid \mathcal{Z}_t^-\right]$, for $k = 0, 1, \ldots, \infty$, i.e.

$$\mathcal{X}_t^{+/-} = \bigvee_{k=0}^{\infty} E_{\|\mathcal{Z}_{[t,t+k)}}\left[\mathcal{Y}_{t+k} \mid \mathcal{Z}_t^-\right]. \qquad (2.5)$$

The closed vector sum can be terminated at any $k \geq n$ where n is the system order, i.e. the dimension of $\mathcal{X}_t^{+/-}$, in which case it is only required that $\mathcal{Z}_{[t,t+k)} \cap \mathcal{Z}_t^- = \{0\}$.

We would like to stress that the recipe (2.5) just requires computing oblique projections of future outputs ($\mathbf{y}(t+k)$) along the future input and output space ($\mathcal{Z}_{[t,t+k)}$) onto the past data ($\mathcal{Z}_t^-$). This yields the following procedure (based on infinite past data) to estimate the system matrices (A, B, C):

(1) Compute the oblique projections

$$E_{\|\mathcal{Z}_{[t,t+k)}}\left[\mathcal{Y}_{t+k} \mid \mathcal{Z}_t^-\right] \quad , \quad k = 0, \ldots, K \qquad (2.6)$$

and find a "best" n-dimensional [4] basis $\mathbf{x}(t)$ for the subspace $\mathcal{X}_t^{+/-}$ generated by these oblique predictors.

(2) Repeat the same procedure shifting time to $t+1$, to get a (coherent) basis in $\mathcal{X}_{t+1}^{+/-}$.

(3) Solve by standard least squares the regression (2.1) for the system matrices (A, B, C).

[3] Recall that strict stability of the predictor is always required for prediction error methods, and it is also postulated in (Jansson 2003).

[4] Here the system order n is also assumed to be known. Of course any consistent order estimation procedure used in subspace identification would do. Order estimation is performed in most subspace identification algorithms by a (weighted) SVD truncation step which shall not discuss in this paper.

This algorithm will be further discussed in Section 3.

From the above it takes a very short step to realize that $\mathcal{X}_t^{+/-}$ is also the state space of the *output predictor* $\hat{\mathbf{y}}(t|t-1)$ based on the joint past data. Actually, since $\hat{\mathbf{y}}(t|t-1) = \mathbf{y}(t) - \mathbf{e}(t)$, the latter is described essentially by the same model as the whitening filter (2.3), namely

$$\begin{cases} \mathbf{x}(t+1) &= \bar{A}\mathbf{x}(t) + B\mathbf{u}(t) + K\mathbf{y}(t) \\ \hat{\mathbf{y}}(t|t-1) &= C\mathbf{x}(t) \end{cases} \qquad (2.7)$$

where $\bar{A} := A - KC$. Hence it follows that the state space of this model, still equal to $\mathcal{X}_t^{+/-}$, must be the oblique predictor space of $\hat{\mathbf{y}}$ given $\mathbf{z}$, namely

$$\mathcal{X}_t^{+/-} := E_{\|\mathcal{Z}_t^+}\left[\hat{\mathcal{Y}}_t^+ \mid \mathcal{Z}_t^-\right]. \qquad (2.8)$$

Therefore *formulas* (2.2), (2.4), (2.5) *and* (2.8) *turn out to define the same object*. In conclusion, exactly as prediction error methods are based on the identification of the predictor $\hat{\mathbf{y}}_\theta(t|t-1)$, subspace methods can be based on the identification of the state space $\mathcal{X}_t^{+/-}$ of the predictor realization (2.7).

Even though at a superficial look the algorithm of (Chiuso and Picci 2005) described by the steps (1), (2), (3) above might look different, it is actually an implementation of formula (2.8) for the construction of the state space. The sequential steps required to implement (2.6) for $k = 1, 2, \ldots$, are in fact hidden in the computation of the predictors $\hat{\mathbf{y}}(s|s-1)$, $s \in [t, T)$. A formal proof of this fact follows from the identities

$$\begin{aligned} E_{\|\mathcal{Z}_t^+}\left[\hat{\mathbf{y}}(s|s-1) \mid \mathcal{Z}_t^-\right] &= E_{\|\mathcal{Z}_{[t,s)}}\left[\hat{\mathbf{y}}(s|s-1) \mid \mathcal{Z}_t^-\right] \\ &= E_{\|\mathcal{Z}_{[t,s)}}\left[\mathbf{y}(s) \mid \mathcal{Z}_t^-\right] \end{aligned}$$

since $\hat{\mathbf{y}}(s|s-1) \in \mathcal{Z}_s^-$, $s \in [t, T)$.

It is remarkable that it has taken over a decade to realize that the predictor model could lead to a more general and robust method for subspace identification which works equally well in the presence of feedback.

3. SUBSPACE IDENTIFICATION WITH FINITE DATA

In the previous discussion it has been established that the predictor space $\mathcal{X}_t^{+/-}$ can be computed without bias (and hence the system matrices can be consistently estimated from (2.1)) in case *infinite past data* were available. However it has recently been shown by (Chiuso and Picci 2005) that using closed-loop data on a *finite* past horizon $t - t_0$, to compute the predictors (which is the only realistic way to go in practice) may lead to biased estimates. Closed-loop subspace procedures such as (Qin and Ljung 2003, Jansson 2003, Chiuso and Picci 2005) may fail to give consistent estimates unless some particular condition on the zeros of

the joint spectrum are satisfied. The reason is that the *transient* predictor (i.e. the transient Kalman filter based on finite past data) requires also modeling of $\mathbf{u}$. In particular a minimal realization of the transient predictor has in general larger dimension than the innovations model (2.1) [5] while, when feedback is absent, the input process $\mathbf{u}$ can be considered exogenous and modeling of $\mathbf{u}$ can be completely avoided. In this case the finite history transient predictor defined as $\hat{\mathbf{y}}_{t_0}(t \mid t-1) := E\left[\mathbf{y}(t) \mid \mathcal{Z}_{[t_0,t)} \vee \mathcal{U}_{[t,T]}\right]$, admits a state space realization of the same dimension of (2.1), which, remarkably, involves just the stationary parameters [6] (A, B, C) (Van Overschee and De Moor 1994, Chiuso and Picci 2004*b*). This is essentially due to the fact that when data are collected in open loop, the effect of future inputs can always be encoded in the initial condition $\hat{\mathbf{x}}(t_0) := E[\mathbf{x}(t_0) \mid \mathcal{U}_{[t_0,T]}]$ (see (Van Overschee and De Moor 1994, Chiuso and Picci 2004*b*)) while in closed loop this is no longer possible. This is precisely the point where feedback plays a crucial role. To put it another way:

(1) In *open loop*, the future input signal after t_0 is independent of the initial condition $\mathbf{x}(t_0)$ and of the whole innovation process $\mathbf{e}$.
(2) In *closed loop*, inputs are generated through feedback and depend on the initial condition *and* on the innovation sequence.

It follows that in feedback systems, measuring the input helps the estimation of the initial condition so one should not ignore the input dynamics. The bias is due to the fact that using the input dynamics (which we instead disregard) the initial condition can be estimated more accurately. This effect disappears for finite-memory systems (i.e. of the ARX type), see (Chiuso and Picci 2005) for details.

At this point one might wonder if classical prediction error methods also resent of finite regression horizon effects or if, under the same circumstances, they behave intrinsically better (i.e. do provide consistent estimates) than subspace methods. Note that, in order to make the optimization problem tractable, prediction error methods are generally based on the stationary predictor (this is the reason why, for finite data length, PEM *are not* equivalent to maximum likelihood even with Gaussian innovations). However, the difference between the log-likelihood and the average prediction error based on the stationary predictor becomes negligible as the data length goes to infinity. In fact in PEM a time average of some norm of the (approximate) prediction errors $\mathbf{y}(t) - \hat{\mathbf{y}}_\theta(t|t-1)$, $t = 1, .., N$ each based on data from 0 to $t-1$, is minimized. As the data length N tends to infinity the effect of the terms with "wrong" initial condition on the predictors is $O(1/N)$ and can be neglected asymptotically in the optimization. On the other hand, in subspace methods, unless the "past horizon" $t - t_0$ is made to grow logarithmically with N (see for instance (Bauer and Ljung 2001)), some bias will be present. It should however be kept in mind that the computational complexity of subspace algorithms grows significantly when enlarging the past horizon and the problem becomes even more relevant for MIMO systems, which are the principal application area of subspace methods.

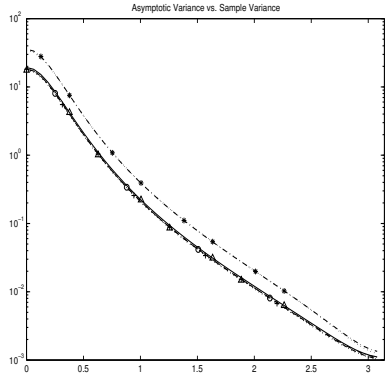

Fig. 1. Asymptotic Variance (Monte Carlo estimate) vs. normalized frequency ($\omega \in [0, \pi]$) Solid with triangles ($\triangle$) PEM, dashed with stars stars ($*$): "innovation estimation", dashed with crosses ($+$) "predictor-based" algorithm, dashed with circles (o): Jansson's algorithm, dotted with crosses ($+$): asymptotic variance for "predictor-based", dotted with stars ($*$): asymptotic variance for "innovation estimation".

4. SIMULATION RESULTS

In this section we report some simulation results. We consider the first order ARMAX model

$$\begin{aligned} \mathbf{x}(t+1) &= 0.7\mathbf{x}(t) + \mathbf{u}(t) + \mathbf{e}(t) \\ \mathbf{y}(t) &= \mathbf{x}(t) + \mathbf{e}(t) \end{aligned}$$

driven by the input process (note there is feedback)

$$\mathbf{u}(t) = 5\mathbf{n}(t) - 1.5\mathbf{y}(t)$$

where $\mathbf{n}(t)$ and $\mathbf{e}(t)$ are uncorrelated unit variance white noises. We compare the Monte Carlo estimate (over 500 trials with $N = 1000$ data points each) of the transfer function variance (normalized by N) of several algorithms (see figure description) with the asymptotic variance formulas obtained in (Chiuso 2004) and the Cramér

[5] The general structure of finite-interval predictors in the presence of feedback has been studied in (Chiuso and Picci 2005) where the explicit expressions are reported.
[6] Here we are interested only in the "deterministic" transfer function $F(z)$. Of course different considerations hold for the Kalman gain K which requires the solution of a Riccati equation. .

Rao lower bound. Both the algorithm based on the predictor and Jansson's approach are indistinguishable from PEM.

5. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper we have shown that with an infinite amount of data (esp. with infinite past data), subspace identification of feedback systems could be implemented successfully, based on the idea of *predictor model identification*. In the more realistic case of a finite past horizon $t-t_0$, the estimates turn out to be generally biased, the amount of bias decreasing as the length of chosen past horizon increases. Computations based on the asymptotic variance expressions found in (Chiuso 2004) show that neither the "predictor based algorithm" nor the "whitening filter" algorithm are efficient in general. Future work will address these aspects.

REFERENCES

Bauer, D. and L. Ljung (2001). Some facts about the choice of the weighting matrices in larimore type of subspace algorithm. *Automatica* **36**, 763–773.

Caines, P.E. and C.W. Chan (1976). Estimation, identification and feedback. In: *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, Eds.). Academic. pp. 349–405.

Chiuso, A. (2004). Asymptotic variance of closed-loop identification algorithms. *Submitted to IEEE Trans. on Aut. Control*. available at http://www.dei.unipd.it/∼chiuso.

Chiuso, A. (2005). On the relation between cca and predictor based subspace identification. submitted to CDC-ECC 2005, available at http://www.dei.unipd.it/∼chiuso.

Chiuso, A. and G. Picci (2003). *Geometry of Oblique Splitting, Minimality and Hankel Operators*. pp. 85–124. Number 286 In: *Lect. Notes in Control and Information Sciences*. Springer.

Chiuso, A. and G. Picci (2004*a*). Numerical conditioning and asymptotic variance of subspace estimates. *Automatica* **40**(4), pp. 677–683.

Chiuso, A. and G. Picci (2004*b*). On the ill-conditioning of subspace identification with inputs. *Automatica* **40**(4), pp. 575–589.

Chiuso, A. and G. Picci (2005). Consistency analysis of some cloded-loop subspace identification methods. *Automatica* **41**(3), 377–391.

Chou, C.T. and M. Verhaegen (1997). Subspace algorithms for the identification of multivariable dynamic error-in-variables models. *Automatica* **33**(10), 1857–1869.

Forsell, U. and L. Ljung (1999). Closed loop identification revisited. *Automatica* **35**, 1215–1242.

Gevers, M.R. and B.D.O. Anderson (1981). Representation of jointly stationary feedback free processes. *Intern. Journal of Control* **33**, 777–809.

Granger, C.W.J. (1963). Economic processes involving feedback. *Information and Control* **6**, 28–48.

Hannan, E.J. and D.S. Poskitt (1988). Unit canonical correlations between future and past. *The Annals of Statistics* **16**, 784–790.

Jansson, M. (2003). Subspace identification and arx modelling. In: *Proceedings of SYSID 2003*. Rotterdam.

Larimore, W.E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In: *Proc. 29th IEEE Conf. Decision & Control*. Honolulu. pp. 596–604.

Lindquist, A. and G. Picci (1996). Canonical correlation analysis, approximate covariance extension and identification of stationary time series. *Automatica* **32**, 709–733.

Ljung, L. (1997). *System Identification; Theory for the User*. Prentice Hall.

Ljung, L. and T. McKelvey (1996). Subspace identification from closed loop data. *Signal Processing* **52**(2), 209–216.

Ng, T.S., G.C. Goodwin and B.D.O. Anderson (1977). Identifiability of MIMO linear dynamic systems operating in closed loop. *Automatica* **13**, 4776–485.

Picci, G. (1997). Oblique splitting susbspaces and stochastic realization with inputs. In: *Operators, Systems and Linear Algebra* (D. Prätzel-Wolters U. Helmke and E. Zerz, Eds.). pp. 157–174. Teubner,. Stuttgart.

Picci, G. and T. Katayama (1996). Stochastic realization with exogenous inputs and "subspace methods" identification. *Signal Processing* **52**, 145–160.

Qin, S.J. and L. Ljung (2003). Closed-loop subspace identification with innovation estimation. In: *Proceedings of SYSID 2003*. Rotterdam.

Van Overschee, P. and B. De Moor (1993). Subspace algorithms for the stochastic identification problem. *Automatica* **29**, 649–660.

Van Overschee, P. and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic– stochastic systems. *Automatica* **30**, 75–93.

Van Overschee, P. and B. De Moor (1997). Closed loop subspace systems identification. In: *Proceedings of 36th IEEE Conference on Decision & Control*. San Diego, CA. pp. pp. 1848–1853.

Verhaegen, M. (1994). Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica* **30**, 61–74.