VALIDATION TEST BASED PARAMETER UNCERTAINTY VERSUS ANALYSIS-BASED CONFIDENCE BOUNDS

Sippe G. Douma * **Xavier J.A. Bombois** * **Paul M.J. Van den Hof** *

* *Delft Center for Systems and Control,*
Delft University of Technology, Mekelweg 2, 2628 CD Delft,
The Netherlands,
s.g.douma@dcsc.tudelft.nl x.j.a.bombois@dcsc.tudelft.nl
p.m.j.vandenhof@dcsc.tudelft.nl

Abstract: Standard Instrumental Variable system identification methods provide for a particular parameter confidence region under a Gaussian distribution. Alternatively, a parameter bounding technique based on a set of constraints on a cross-correlation function directly provides for a parameter uncertainty set. This paper relates the two uncertainty regions under the standard assumption of additive noise on the measured data, both in case undermodelling is considered and in case it is not. The ellipsoidal region associated with the IV estimation technique is strongly related to the polytopic region induced by the cross-correlation constraints as both techniques are based on the same set of regressors. However, they differ due to the fact that the former incorporates a covariance between errors, while the latter is limited to the variances of the errors only. The results are also discussed in terms of the standard model validation which is identical in nature to the cross-correlation parameter bounding technique.

Keywords: parameter uncertainty, confidence region, set membership, model validation

1. INTRODUCTION

Given measured input and output data of a system any technique can be applied to yield a model describing the relation between the input and the output. However, a model is useful only when some indication is provided of its precision and accuracy. In turn, the measure of confidence in the model depends entirely on the assumption made on the data-generating system.

In standard prediction error identification the output data is assumed to be noise-corrupted by a stationary stochastic sequence with a particular distribution. Based on this assumption a confidence region of estimated parameters can be constructed, both when assuming that the data-generating system is in the model class (e.g. Ljung (1999)) and when undermodelling is taken into consideration (e.g. Hakvoort and Van den Hof (1997)).

Alternatively, parameter bounding methods or set-membership techniques pose a number of deterministic bounds on the noise (e.g. Milanese and Vicino (1991); Veres and Norton (1991)). These techniques result directly in a set of feasible parameters. While standard set membership techniques place a time domain bound on the additive noise, Hakvoort en Van den Hof (1995) have proposed a bound on the cross-correlation between the noise and some instrumental signal. The former methods are known to lead to conservative results, or even to no consistency, as the noise is allowed to take a worst-case realization within the noise bounds. Bounds on the cross-correlation of the noise with a instrumental signal give rise to a consistent parameter bounding method leading to less conservative results.

In this paper the standard prediction error confidence region is related to the set of parameters resulting from the cross-correlation parameter bounding technique. In particular, the IV estimation method without undermodelling (e.g. Ljung (1999); Söderström and

Stoica (1989)) and with undermodelling (Hakvoort and Van den Hof, 1997) is considered with respect to the cross-correlation method of Hakvoort and Van den Hof (1995). Since the cross-correlation method is in nature identical to a posteriori model validation tests used in standard system identification, the results will also be interpreted in light of model validation based on data.

After preliminary information is given in section 2, section 3 introduces the notation used in the paper and briefly discusses the different methods. Based on a particular set of assumptions, the methods are analyzed in section 4. The comparison itself is made in sections 5 and 6 for the most general case and for illustrative special cases, respectively. After a small section with further remarks, the conclusions are presented.

2. PRELIMINARIES

Here we consider discrete linear time-invariant finite dimensional models which, additionally, are linearly parametrized. These parametrized models $G(q,\theta)$ are represented by

$$G(q,\theta) = \sum_{k=0}^{n-1} g(k) F_k(q),$$

where θ denotes the parameter vector $\begin{bmatrix} g(0) & g(1) & \cdots & g(n-1) \end{bmatrix}^T \in \mathbb{R}^n$ and $F_k(q)$ the basis functions. A standard choice for $F_k(q)$ is to consider orthonormal basis functions (see e.g. Heuberger *et al.* (1995)). This class contains the well-known FIR, Laguerre and Kautz models. For example, for a FIR-model $F_k(q)$ is the standard pulse basis, i.e. $F_k(q) = q^{-k}$.

3. METHODS

3.1 *Method 1: Instrumental Variable estimate*

The procedure of the Instrumental Variable Estimation method is discussed here briefly, as it is documented in many references (e.g. Ljung (1999); Söderström and Stoica (1989)).

Procedure of the IV estimation

1. Choose an instrumental variable signal $r(t)$ (For properties of $r(t)$ see section 4.2).
2. Filter the measured input $u(t)$ and the instrumental variable $r(t)$ with the basis functions $F_k(q)$ to create signal sequences $x_k(t)$ and $z_k(t)$,

   $$x_k(t) = F_k(q)u(t), \qquad z_k(t) = F_k(q)r(t),$$

   and define the column-vectors
   $\phi(t) := \begin{bmatrix} x_0(t) \dots x_{n-1}(t) \end{bmatrix}^T$
   $\zeta(t) := \begin{bmatrix} z_0(t) \dots z_{n-1}(t) \end{bmatrix}^T$.
3. Obtain the instrumental variable estimate $\hat{\theta}_N^{IV}$ by

   $$\hat{\theta}_N^{IV} := \underset{\theta\in\Theta}{sol} \left\{ \frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)\varepsilon(t,\theta) = 0 \right\}, \quad (1)$$

   with $\Theta \subset \mathbb{R}^n$ an appropriate parameter space, $t_s > 0$ representing the starting sample used in the estimation, and $\tilde{N} = N - t_s + 1$. The term $\varepsilon(t,\theta)$ denotes the residue-signal

   $$\begin{aligned} \varepsilon(t,\theta) &= y(t) - G(q,\theta)u(t) \\ &= y(t) - \phi^T(t)\theta. \end{aligned} \quad (2)$$

Straightforward algebraic manipulation reveals the solution to (1) as

$$\hat{\theta}_N^{IV} = \left[\frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)\phi^T(t) \right]^{-1} \frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)y(t). \quad (3)$$

provided that the matrix to be inverted is nonsingular (which depends on the choice of $r(t)$). It is to be noted that when the input $u(t)$ itself is taken as instrumental variable $r(t)$, expression 3 gives the standard linear least squares estimate of the expansion coefficients.

Parameter confidence region of the IV estimate

A confidence region $\mathcal{D}^{IV}$ of the instrumental variable estimate is usually constructed based on the assumption that $\hat{\theta}_N^{IV}$ is Gaussian distributed with the true parameter vector θ_0 as mean (Ljung, 1999). In section 4 the underlying assumptions of this approach are discussed.

Theoretically, every estimate $\hat{\theta}_N^{IV} \in \mathcal{N}(\theta_0, P_\theta)$ is contained, with a probability of α, within the standard ellipsoidal region associated with the Gaussian distribution defined by all θ such that

$$(\theta - \theta_0)^T P_\theta^{-1} (\theta - \theta_0) \le c_{\chi,\alpha}, \quad (4)$$

where $c_{\chi,\alpha}$ corresponds to a probability α in the chi-squared distribution with n degrees of freedom. This follows from the fact that (the spectral factor of) the inverse of the matrix P_θ transforms the parameters to a set of uncorrelated Gausian distributed parameter each having a standard deviation of 1. The norm of this transformed parameter vector then has a chi-squared distribution with n degrees of freedom. The shape of the ellipsoid is defined by the square root of the eigenvalues of P_θ and its orientation by the associated eigenvectors.

Using (4), a confidence region $\mathcal{D}^{IV}$ containing the unknown true parameter vector θ_0 at a certain probability level α can be constructed with a particular estimate $\hat{\theta}_N^{IV}$ and based on an estimation $\hat{P}_\theta$ of the covariance matrix as

$$\mathcal{D}^{IV} = \left\{ \theta \mid \left(\theta - \hat{\theta}_N^{IV}\right)^T \hat{P}_\theta^{-1} \left(\theta - \hat{\theta}_N^{IV}\right) \le c_\alpha \right\}. \quad (5)$$

Note that the the probability level α corresponding to c_α now depends on the properties of the estimate $\hat{P}_\theta$ as well (Ljung, 1999).

3.2 *Method 2: Cross-correlation method*

Parameter bounding identification methods (e.g. Milanese and Vicino (1991); Veres and Norton (1991)) are based on a set of linear constraints, e.g. constraints on the noise signal or a cross-correlation function. Here we consider the procedure proposed in Hakvoort and Van den Hof (1995):

Procedure of the cross-correlation method

1-2 as in the IV method

3. Define the feasible *set* of parameters by

$$\mathcal{D}^{cc} = \left\{ \theta \mid b_l(k) \leq \frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta_k(t)\varepsilon(t,\theta) \leq b_u(k) \quad , \right.$$
$$\left. k = 1, \ldots, n. \right\} \tag{6}$$

This set of parameters defines a polytope in the parameter space $\mathbb{R}^n$ as $\varepsilon(t,\theta)$ is linear in the parameters. This polytope is defined by $2n$ half planes in $\mathbb{R}^n$ of which the n halfplanes induced by the lower bounds b_l are parallel to those induced by the upper bounds b_u. The corner points of this polytope are solutions to a convex optimization problem (e.g. Broman and Shensa (1990)). However, note that in (6) the number of constraints equals the number of parameters. In this case a simple analytical solution exists, which is formulated in the following proposition:

Proposition 1. The solution space $\mathcal{D}^{cc} \in \mathbb{R}^n$ of all parameter vectors θ satisfying the cross-correlation constraint of (6) defines a polytope with 2^n corner points induced by the solutions to

$$\frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)\varepsilon(t,\theta) = Be_j \quad , j = 1, \ldots, 2^n.$$

where the vector e_j denotes the j-th Euclidean vector taking the j-the column from the matrix $B \in \mathbb{R}^{n \times 2^n}$ containing all possible combination of $b_l(k)$ and $b_u(k)$. For example, when $n = 3$, $B =$

$$\begin{bmatrix} b_l(0) & b_l(0) & b_l(0) & b_l(0) & b_u(0) & b_u(0) & b_u(0) & b_u(0) \\ b_l(1) & b_l(1) & b_u(1) & b_u(1) & b_u(1) & b_l(1) & b_u(1) & b_l(1) \\ b_l(2) & b_u(2) & b_l(2) & b_u(2) & b_u(2) & b_u(2) & b_l(2) & b_l(2) \end{bmatrix}$$

In other words, B simply defines the corner points of a n-dimensional box.
In particular, with $R = \frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)\phi^T(t)$, the corner points of $\mathcal{D}^{cc}$ are given by the vectors

$$\theta_j^{cc} = R^{-1} \frac{1}{\tilde{N}} \sum_{t=t_s}^{N} \zeta(t)y(t) - R^{-1}Be_j, \tag{7}$$

for $j = 1, \ldots, 2^n$.
Proof is omitted as it is straightforward. □

Confidence region of cross-correlation method
The properties of the polytope defined by the solution set $\mathcal{D}^{cc}$ of (6) can be discerned from the operation on

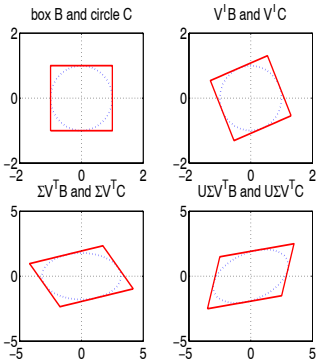

Fig. 1. *Transformation of a box B and a circle C induced by $R^{-1} = U_R\Sigma_R V_R^T$.*

the 'box' B by the matrix R^{-1}. Consider the singular value decomposition $U_R\Sigma_R V_R^T$ of R^{-1}. The matrix R^{-1} is then seen to rotate the 'box' B by V_R^T, dilate the result along the Euclidean axes with the elements of Σ_R, after which the matrix U_R induces a second rotation (see figure 1 for $n = 2$).

4. ASSUMPTIONS AND ASSOCIATED ANALYSIS

Expressions (3) and (7) show that:

$$\theta_j^{cc} = \hat{\theta}_N^{IV} - R^{-1}Be_j. \tag{8}$$

In other words, the nominal parameter vector of the cross–correlation based set $\mathcal{D}^{cc}$ equals the IV estimate $\hat{\theta}_N^{IV}$. The term B represents a confidence region: the possible deviations around the nominal estimate. In section 5 this confidence region $\mathcal{D}^{cc}$ will be compared with the confidence region $\mathcal{D}^{IV}$ of the IV-estimate $\hat{\theta}_N^{IV}$. But first the confidence regions $\mathcal{D}^{IV}$ and $\mathcal{D}^{cc}$ are analyzed in detail.

4.1 *Assumptions*

Measures of confidence in estimated models depend almost entirely on a priori assumptions. Here we consider the following assumptions on the data-generating system, based on Hakvoort and Van den Hof (1997):

$$y(t) = \tilde{G}_0(q)u(t) + \bar{G}_0(q)u(t) + G_0(q)u^-(t) + v(t),$$

where

i. $\tilde{G}_0(q)u(t) + \bar{G}_0(q)u(t)$, denotes the part of the output $y(t)$ directly related to the measured sequence u by means of a linear-time invariant operator $G_0(q)$, decomposed into a modelled ($\tilde{G}_0(q)$) and unmodelled part ($\bar{G}_0(q)$). That is,

$$G_0(q)u(t) = \tilde{G}_0(q)u(t) + \bar{G}_0(q)u(t)$$
$$= \phi^T(t)\theta_0 + \sum_{k=n+1}^{\infty} g_0(k)F_k(q)u(t),$$

($u(t) := 0$ outside the measurement interval $t = 1, \ldots, N$)

ii. $G_0(q)u^-(t)$ denotes the effect of initial conditions, i.e. the effect of a sequence u^- preceding the measured $u(t)$ ($u^-(t) := 0$ for $t > 0$) and

iii. $v(t)$ denotes a stochastic component of which only the ensemble properties are constant in time. Here we consider $v(t)$ to be uncorrelated with the sequences u and r.

4.2 *Analysis of the Instrumental Variable estimate*

With the assumptions given above the instrumental variable estimate (3) can be analyzed as

$$\hat{\theta}_N^{IV} = \theta_0 + R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)\gamma(t). \tag{9}$$

with $\gamma(t) = \bar{G}_0(q)u(t) + G_0(q)u^-(t) + v(t)$.
A confidence region around the IV estimate $\hat{\theta}_N^{IV}$ will be based on the error term $R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)\gamma(t)$.
In the following two possible assumptions on this error term are evaluated.

4.2.1. *System is in the model class* In case the true system $G_0(q)$ is considered to be in the model class, i.e. $\bar{G}_0(q)u(t) = 0$ and $u^-(t) = 0$ for all t, expression (9) becomes

$$\hat{\theta}_N^{IV} = \theta_0 + R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t) \tag{10}$$

Using the assumptions on $v(t)$, it follows that

$$\hat{\theta}_N^{IV} \in \mathcal{N}(\theta_0, P_\theta) \tag{11}$$

since $R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t)$ is obtained from linear combinations of $v(t)$. If $v(t)$ is assumed to have a distribution other than Gaussian, the term $R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t)$ will be Gaussian distributed only asymptotically as a result of the Central Limit Theorem. The covariance matrix P_θ is given by

$$P_\theta = R^{-1}\Lambda^N\left(R^{-1}\right)^T \tag{12}$$

where

$$\Lambda^N = E\left[\left(\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t)\right)\left(\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t)\right)^T\right].$$

The covariance matrix P_θ is seen to consists of (co)variances of the term $\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t)$, for which an exact expression is available (Hakvoort and Van den Hof, 1997):

$$\Lambda^N(i,j) = \frac{1}{\tilde{N}}\sum_{\tau=-N+t_s}^{N-t_s} R^N_{\zeta_i\zeta_j}(\tau)R_v(\tau), \tag{13}$$

where

$$R^N_{\zeta_i\zeta_j}(\tau) := \frac{1}{\tilde{N}}\sum_{t=t_s}^{N+\tau}\zeta_i(t)\zeta_j(t-\tau),\ \tau = [-N+t_s, 0]$$

$$R^N_{\zeta_i\zeta_j}(\tau) := \frac{1}{\tilde{N}}\sum_{t=t_s}^{N-\tau}\zeta_i(t+\tau)\zeta_j(t)\ ,\ \tau = [1, N-t_s]$$

The confidence region $\mathcal{D}^{IV}$ of expression (5) is obtained with the estimate of P_θ which follows from applying expressions (12) and (13) with an estimate $\hat{R}_v(\tau)$ of the noise properties.

4.2.2. *Undermodelling* The effects of undermodelling reflected in $\bar{G}_0(q)u(t)$ and of initial conditions reflected in $G_0(q)u^-(t)$ in expression (9) introduce a deterministic but unknown bias in the estimator $\hat{\theta}_N^{IV}$. This bias Δ can be incorporated by means of bounds on its absolute value. That is, it holds that

$$\hat{\theta}_N^{IV} \in \Delta + \mathcal{N}(\theta_0, P_\theta), \quad |\Delta| < \beta,$$

where $\beta = \beta_1 + \beta_2$ and

$$\left|R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)\bar{G}_0(q)u(t)\right| < \beta_1$$

$$\left|R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)G_0(q)u^-(t)\right| < \beta_2.$$

The calculation of the bounds is not elaborated here as they follow easily from adapting the (frequency domain) bounds proposed in Hakvoort and Van den Hof (1997). A confidence region in the parameter space is given by the union of ellipsoids of expression (5) and section (4.2.1) centred at all possible mean values $\hat{\theta}_N^{IV} + \Delta$, with $|\Delta| < \beta$. It can be said that this union of ellipsoids contains the true system with a probability *larger than* α.

Alternatively, the effect of initial conditions could be considered as a stochastic influence similar to $v(t)$ as it depends on the unknown $u^-(t)$. Though not elaborated here, in that case, the covariance matrix P_θ could be extended additively by a matrix containing the covariances of the term $R^{-1}\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)G_0(q)u^-(t)$. The incorporation of the undermodelling would proceed as above with $\beta = \beta_1$.

4.3 *Analysis of the cross-correlation method*

The IV-estimation method provides for a nominal estimate based on data and a confidence region based on assumptions. The cross-correlation method, on the other hand, directly provides for a confidence region $\mathcal{D}^{cc}$ based on a data set. However, the cross-correlation method is not without assumptions. The bounds used in the cross-correlation method have to be chosen based on assumptions on the underlying system. Or, alternatively, the bounds chosen in the cross-correlation method have to be interpreted with

respect to assumptions if statements are to be made about the quality of the resulting parameter set.

Considering the assumptions of section 4.1, the cross-correlation term $\frac{1}{N}\sum_{t=t_s}^{N}\zeta(t)\varepsilon(t,\theta)$ in (6) is evaluated as

$$\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)\varepsilon(t,\theta) = R\left(\theta_0-\theta\right)+\frac{1}{\tilde{N}}\sum_{t=t_s}^{N}\zeta(t)v(t),$$

in case the system is assumed to be in the model class. Clearly, for $\mathcal{D}_{cc}$ to contain the true parameter vector θ_0, the bounds $b_l(k)$ and $b_u(k)$ should allow for the error term $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$. They could be interpreted as (deterministic) upper bounds on the amplitude of $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$, or, based on a stochastic noise assumption, as bounds on the standard deviation of $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$ times a factor $\sqrt{c_\alpha}$ corresponding to a certain level of probability. That is, with $b_l(k) = -b_u(k) = b(k)$,

$$b(k) = \sqrt{c_\alpha \Lambda^N(k,k)}.$$

Similarly, in case of undermodelling, the bounds $b_l(k)$ and $b_u(k)$ should explain the effects of both the noise term $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$ and the undermodelling, i.e.

$$b(k) = \sqrt{c_\alpha \Lambda^N(k,k)} + \beta,$$

where Λ^N and β are as in section 4.2.

5. COMPARISON OF THE CONFIDENCE REGIONS

From the discussion in the previous section and in particular from expressions (5), (12) and (7) it is seen that

$$\mathcal{D}^{IV} = \left\{\theta \mid \theta = \hat{\theta}_N^{IV} + \Delta_\theta\,, \Delta_\theta \in R^{-1}\mathcal{E}\left(c_\alpha\hat{\Lambda}^N\right)\right\}$$
$$\mathcal{D}^{cc} = \left\{\theta \mid \theta = \hat{\theta}_N^{IV} + \Delta_\theta\,, \Delta_\theta \in R^{-1}\mathcal{B}(B)\right\}, \quad (14)$$

where $\mathcal{E}(A)$ denotes the ellipsoid defined by all $x \in \mathbb{R}^{n\times 1}$ such that $x^T A^{-1} x < 1$ and $\mathcal{B}(B)$ denotes the interior of the n-dimensional box with corner points defined in the matrix $B \in \mathbb{R}^{n\times 2^n}$. That is, the polytope of $\mathcal{D}^{cc}$ is formed by the transformation induced by the matrix $R^{-1}$of the *hyperbox* B (cf. section 3.2). Similarly, the ellipsoid of $\mathcal{D}^{IV}$ according to the estimated matrix $\hat{P}_\theta$ can be interpreted as the transformation induced by R^{-1} of an *ellipsoid* defined by $\hat{\Lambda}^N$ based on expression (13) with an estimate $\hat{R}_v(\tau)$ of the noise properties. (see figure 2).

With expression (14) the strong connection between $\mathcal{D}^{cc}$ and $\mathcal{D}^{IV}$ is made clear. For example, the smallest hyperbox $\mathcal{D}^{cc}$ containing the ellipsoid $\mathcal{D}^{IV}$ can easily be found (see figure 2). The link between the two sets results from the fact that the same regressors are used. The essential difference between the two methods lies with the information used on the

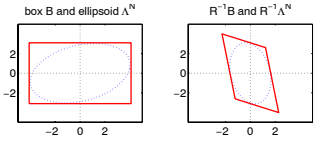

Fig. 2. *Confidence regions for $\mathcal{D}^{cc}$ and $\mathcal{D}^{IV}$: transformation by R^{-1} of a box B and of an ellipsoid associated with Λ^N.*

error term $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$. While in $\mathcal{D}^{cc}$ only the size of the variance at each constraint k can be taken into account, in $\mathcal{D}^{IV}$ also the correlation of the errors $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$ over all n constraints is incorporated by means of the covariance matrix Λ^N.

$\mathcal{D}^{IV}$ and $\mathcal{D}^{cc}$ are confidence regions containing θ_0 at a certain probability level when the true system is in the model class. However, in the presence of undermodelling or initial conditions similar confidence regions as (14) can be found. The size of the ellipsoid $\mathcal{E}\left(c_\alpha\hat{\Lambda}^N\right)$ and the hyperbox $\mathcal{B}(B)$ simply have to be increased as indicated in sections 4.2.2 and 4.3.

In the next section some special cases will be presented further illustrating the link between the two sets.

6. SPECIAL CASES

In some special cases the connection between the ellipsoidal $\mathcal{D}^{IV}$ and the boxed $\mathcal{D}^{cc}$ is particularly straightforward and elucidating.

6.1 Λ^N *is diagonal*

If the errors induced by $\frac{1}{N}\sum_{t=t_s}^{N}\zeta_k(t)v(t)$ are uncorrelated over k, the two methods are both based on variances of the error term only. As such they would use the same information to construct the confidence region. For example, assume $i)$ $\bar{G}_0(q)u(t) = 0$ and $u^-(t) = 0$ for all t, $ii)$ $v(t)$ is white noise, i.e. $R_v(\tau) = 0$ for $\tau \neq 0$ and $iii)$ the signal $\zeta_k(t)$ is such that $R^N_{\zeta_i\zeta_j}(\tau) = 0$ for $i \neq j$. Expressions (13) then shows Λ^N to be diagonal. The third condition could be achieved exactly by a custom made sequence $\zeta_k(t)$, or asymptotically in N by a realization of a stochastic white noise sequence $\zeta_k(t)$ (e.g. $F_k(q) = q^{-k}$ (FIR) and $r(t)$ a white noise sequence).
Figure 3 illustrates how the smallest hyperbox $\mathcal{D}^{cc}$ embedding $\mathcal{D}^{IV}$ is obtained by taking $b_l(k) = -b_u(k) = \sqrt{c_\alpha\lambda_k}$, where the λ_k are the entries (variances) of the now diagonal matrix Λ^N.
Alternatively, the hyperbox $\mathcal{D}^{cc}$ could be chosen to have the same probability level of $\alpha\%$ as the ellipsoidal region $\mathcal{D}^{IV}$. To that end, take $b_l(k) = -b_u(k) = \left(c_{\mathcal{N},\sqrt[n]{\alpha}}\right)\sqrt{\lambda_k}$, where $c_{\mathcal{N},\sqrt[n]{\alpha}}$ corresponds to a probability in a one-dimensional standard Gaussian distribution $\mathcal{N}(0,1)$, such that for $x \in \mathcal{N}(0,1)$ $\text{prob}\left(|x| \leq c_{\mathcal{N},\sqrt[n]{\alpha}}\right) = \sqrt[n]{\alpha}$.

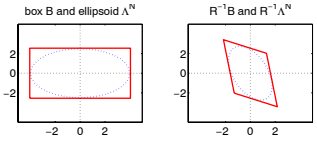

Fig. 3. *Confidence regions for $\mathcal{D}^{cc}$ and $\mathcal{D}^{IV}$ in the special case that Λ^N is diagonal.*

Two parameter case

Another special alignment of $\mathcal{D}^{cc}$ and $\mathcal{D}^{IV}$ occurs in case $i)$ $n = 2$, $ii)$ $\bar{G}_0(q)u(t) = 0$ and $u^-(t) = 0$ for all t, $iii)$ $v(t)$ is white noise, i.e. $R_v(\tau) = \sigma_v^2$ for $\tau = 0$ and zero otherwise, $iii)$ $r = u$ and $iv)$ $F_k(q)$ the FIR-basis. Then $\Lambda^N = \frac{\sigma_v^2}{N} R$ is a 2×2 Toeplitz matrix with $R_{12} = R_{21}$ and $R_{11} = R_{22}$. Due to this special structure, the eigenvectors of R are in the direction of $[1, 1]^T$ and $[1, -1]^T$. Consequently, the ellipsoid associated with Λ^N has its principal axis at 45 and -45 degrees. If B is chosen to define a square box by $b_l(1) = b_l(2) = -b_u(1) = -b_u(2) = b$, the diagonals of this box are aligned with the principal axis of the ellipsoid. Now, the subsequent operation on this square by R^{-1} in expression (14) consists of a rotation again over 45 degrees, a dilation and a subsequent rotation over 45 degrees. Figure 4 illustrates that this results into a rotated rhombus $\mathcal{D}^{cc}$ with its diagonals still aligned with the principal axis of the ellipsoid $\mathcal{D}^{IV}$. The ratio between the diagonals and the principal axis is $\lambda_R 2\sqrt{2}b : \frac{\sigma_v}{\sqrt{N}}\sqrt{\lambda_R}$.

7. SOME REMARKS AND FURTHER RESEARCH

Undermodelling can also be incorporated in the cross-correlation constraints similarly to section 4.2.2. If, on the other hand, the bounds $b(k)$ are chosen based on the properties of $v(t)$ only, the cross-correlation constraints reflect the standard validation test in system identification. From the arguments so far, it is clear that this validation test might suffer from the fact that correlation between the cross-correlation terms is not taken into account. The results also show that any undermodelling effect smaller than the effect of the noise $v(t)$ might not be detected for a particular noise realization. This in particular since the bounds are based on estimated properties of the noise.

8. CONCLUSIONS

Standard Instrumental Variable system identification and a cross-correlation parameter bounding method are strongly related. The ellipsoidal region associated with the IV estimation technique can be tightly linked to the polytopic region induced by the cross-correlation constraints as both techniques are based on the same set of regressors. However, they differ due to the fact that the former incorporates a covariance

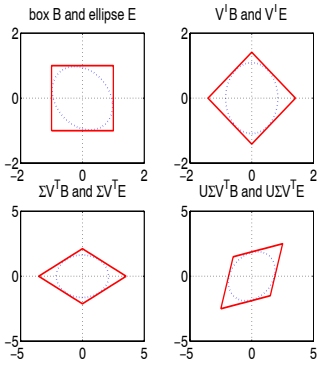

Fig. 4. *Confidence regions for $\mathcal{D}^{cc}$ and $\mathcal{D}^{IV}$ in the special case that $u(t) = r(t)$, $n = 2$ and Λ^N is diagonal.*

between errors, while the latter is limited to the variances of the errors only. Both methods can incorporate undermodelling effects by enlarging their confidence regions with upper bounds on the bias effects. The results hold asymptotically for ARX models too. The insights into the connection between the two methods also apply for the standard practice of model validation.

9. REFERENCES

Broman, V. and M.J. Shensa (1990). A compact algorithm for the intersection and approximation of n-dimensional polytopes. *Mathematics and Computers in Simulation* **32**, 469–480.

Hakvoort, R.G. and P.M.J. Van den Hof (1995). Consistent parameter bounding identification for linearly parametrized model sets. *Automatica* **31**(7), 957–969.

Hakvoort, R.G. and P.M.J. Van den Hof (1997). Identification of probabilistic uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Autom. Control* **42**(11), 1516–1528.

Heuberger, P.S.C., P.M.J. Van den Hof and O.H. Bosgra (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Trans. Autom. Control* **40**, 451–465.

Ljung, L. (1999). *System Identification: Theory for the User (2nd ed.)*. Prentice-Hall, Upper Saddle River, New Jersey, USA.

Milanese, M. and A. Vicino (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica* **27**, 997–1009.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall, Hemel Hempstead, UK.

Veres, S.M. and J.P. Norton (1991). Structure selection for bounded parameter models: consistency conditions and selection criterium. *IEEE Trans. Autom. Control* **36**(7), 474–481.