Delft University of Technology
Delft Center for Systems and Control

Technical report 09-018

An IV-based iterative linear regression algorithm with optimal output error properties

P.M.J. Van den Hof and S.G. Douma

4 November 2008

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
Phone: +31-15-27 8xxxx
Fax: +31-15-27 86679
http://www.dcsc.tudelft.nl

An IV-based iterative linear regression algorithm with optimal Output Error properties [1]

Paul M.J. Van den Hof[*,2] and **Sippe G. Douma**[3]

[*] *Delft Center for Systems and Control,*
Delft University of Technology,
Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract: Identification of output error models generally requires the solution of a nonlinear optimization problem. Both in time domain (Steiglitz and McBride) and in frequency domain (Sanathanan and Koerner) there are alternatives available that try to approach the output error optimum through a sequence of iterative linear regression steps. These algorithms are known to converge to optimal Output Error estimates only under restrictive conditions, i.e. system is in the model set and output noise is white. In this paper an alternative iterative linear regression algorithm is formulated that upon convergence arrives at a (local) optimum of the quadratic cost function of an Output Error model. The alternative algorithm is an iterative instrumental variable algorithm, where the instruments are iteratively updated. In the literature it is known as the simplified refined instrumental variable method (SRIV); however the mentioned optimality properties do not seem to have been formulated explicitly before. The method applies to both time domain and frequency domain identification.

Keywords: system identification, output error models, instrumental variables, iterative algorithms, frequency domain

1. INTRODUCTION

In prediction error identification several attempts have been made to identify accurate linear plant (input-output) models, by simple linear algorithms, and without the need to identify full noise models. This problem is relevant e.g. in situations where the dimensions of the plant limits the possibilities for "crude" nonlinear optimization, or when one is primarily interested in finding good initial model estimates for subsequent nonlinear optimization. In these situations there may be good arguments to focus on plant models only, while arriving at accurate noise models may be postponed to a second stage in the identification procedure.

Examples of model structures and algorithms that are used in these kind of circumstances are the FIR model structure (Ljung (1999)), orthonormal basis functions (ORTFIR) model structures (Heuberger et al. (2005)), and the ARX model structure. The latter option is e.g. often applied in process industry using a high model order in order to avoid model bias due to poor noise modelling, after which model reduction is applied to return to realistic model orders (Zhu (2001, 2006)). An alternative option to deal with the formulated situation is to apply instrumental variable methods, see e.g. Söderström and Stoica (1983).

There exist classical algorithms that attempt to minimize a quadratic output error criterion by successive application of linear regression steps. Next to pseudo-linear regression algorithms (Ljung (1999)), this concerns an iterative algorithm of subsequent ARX modelling and regression filtering, referred to as the method of Steiglitz and McBride, (Steiglitz and McBride (1965)). In the frequency domain an equivalent approach is referred to as the method of Sanathanan and Koerner (Sanathanan and Koerner (1963)) that is often applied in frequency domain identification problems, see e.g. Bayard (1994) and de Callafon et al. (1996). The basic properties of these iterative algorithms is that convergence to an output error optimal model is only known to exist under idealistic conditions, such as white (independent) output noise on the data, while no properties are known for situations of approximate modelling (system does not belong to the model set).

In this paper we will analyze an alternative iterative linear regression method that upon convergence can be shown to have optimal output error properties. The algorithm falls into the class of iterative IV algorithms, and has been presented before, see e.g. Young (1976); Young et al. (2008). However it seems that its optimality properties have not been specifically formulated earlier, or at least have not been recognized as such.

After an introduction of the standard identification setting and the output error identification problem in the time domain, we focus on iterative methods to solve the output error problem. Next we show the relation with IV

[1] 4 November 2008, revised 19 February 2009.
[2] Corresponding author; p.m.j.vandenhof@tudelft.nl
[3] Sippe Douma is now with Shell International Exploration and Production, Rijswijk, The Netherlands.

estimators and we extend the presented algorithm to be applicable in the frequency domain also.

2. IDENTIFICATION SETTING

We consider a discrete-time input output system given by

$$y(t) = G_0(q)u(t) + H_0(q)e(t) \tag{1}$$

with e a sequence of identically distributed independent random variables (white noise) with variance σ_e^2, and G_0 and H_0 transfer functions of linear time-invariant systems. The input u and output y are measurable.

A predictor model is composed as a collection of two transfer functions $G(q,\theta)$, $H(q,\theta)$ while the one-step predictor is given by Ljung (1999)

$$\begin{aligned}\hat{y}(t|t-1;\theta) = &[H^{-1}(q,\theta) - 1]y(t) + \\ &-H(q,\theta)^{-1}G(q,\theta)u(t)\end{aligned} \tag{2}$$

and the prediction error

$$\varepsilon(t,\theta) = H^{-1}(q,\theta)[y(t) - G(q,\theta)u(t)].$$

The models are parametrized by a parameter vector $\theta \in \mathbb{R}^d$ that is estimated on the basis of input output data $\{u(t), y(t)\}_{t=1,\cdots N}$, according to a quadratic cost function:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta)$$

through

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta).$$

The predictor derivative is denoted by:

$$\psi(t,\theta) = \frac{\partial}{\partial\theta}\hat{y}(t|t-1;\theta) = -\frac{\partial}{\partial\theta}\varepsilon(t,\theta).$$

3. OUTPUT ERROR IDENTIFICATION

In an Output Error (OE) model structure we consider the one-step ahead predictor

$$\hat{y}(t|t-1;\theta) = \frac{B(q,\theta)}{F(q,\theta)}u(t), \tag{3}$$

in other words, $H(q,\theta) \equiv 1$. The process model is parametrized as

$$\frac{B(q,\theta)}{F(q,\theta)} = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}$$

with $\theta^T = [f_1 \cdots f_{n_f}\ b_1 \cdots b_{n_b}]$.

If $\hat{\theta}_N$ is an Output Error estimate, it necessarily has to satisfy $V_N'(\hat{\theta}_N) = 0$ or equivalently

$$\frac{1}{N}\sum_{t=1}^{N}[y(t) - \frac{B(q,\hat{\theta}_N)}{F(q,\hat{\theta}_N)}u(t)] \cdot \psi(t,\hat{\theta}_N) = 0. \tag{4}$$

By defining

$$y_F(t) = F(q,\hat{\theta}_N)^{-1}y(t); \quad u_F(t) = F(q,\hat{\theta}_N)^{-1}u(t)$$

equation (4) can be rewritten as

$$\frac{1}{N}\sum_{t=1}^{N}[F(q,\hat{\theta}_N)y_F(t) - B(q,\hat{\theta}_N)u_F(t)] \cdot \psi(t,\hat{\theta}_N) = 0. \tag{5}$$

The parameter estimate $\hat{\theta}_N$ satisfying these equations can now be written in a linear regression-type equation. To this end we denote:

$$\varphi_F^T(t,\hat{\theta}_N) = [-y_F(t-1) \cdots - y_F(t-n_f) \cdots \cdots u_F(t-1) \cdots u_F(t-n_b)],$$

being a vector with dimension $n = n_b + n_f$,

$$\Phi_F^T(\hat{\theta}_N) := [\varphi_F(1,\hat{\theta}_N) \cdots \varphi_F(N,\hat{\theta}_N)]$$

and

$$\Psi^T(\hat{\theta}_N) := [\psi(1,\hat{\theta}_N) \cdots \psi(N,\hat{\theta}_N)].$$

It follows that (5) can be written as

$$\frac{1}{N}\sum_{t=1}^{N}\psi(t,\hat{\theta}_N)[y_F(t) - \varphi_F^T(t,\hat{\theta}_N)\hat{\theta}_N] = 0, \tag{6}$$

leading to

$$\hat{\theta}_N = (\Psi^T(\hat{\theta}_N)\Phi_F(\hat{\theta}_N))^{-1}\Psi^T(\hat{\theta}_N)\mathbf{y}_F \tag{7}$$

with $\mathbf{y}_F = [y_F(1) \cdots y_F(N)]^T$.

Note that (7) is an equation that characterizes $\hat{\theta}_N$; however it cannot be used to *calculate* $\hat{\theta}_N$, as the right hand side of the equation is also dependent on $\hat{\theta}_N$. In Douma and Van den Hof (2005) this expression is used primarily to characterize probabilistic parameter uncertainty bounds for the estimated parameters. However as indicated in Douma (2006) this expression can also be used to implement an iterative scheme to find the OE-solution $\hat{\theta}_N$ as a method of handling the non-convex optimization of $V_N(\theta)$ over θ.

4. ITERATIVE METHODS TO SOLVE AN OE IDENTIFICATION PROBLEM.

There are several pseudo-linear regression (iterative) methods intended to arrive at consistent (and efficient) process model parameter estimates, while relying on simple linear computational tools. One of the popular methods is the so-called Steiglitz-McBride iteration Steiglitz and McBride (1965) that applies a linear regression (ARX) identification to input and output data that is prefiltered with the model denominator inverse. In step k of the iteration, parameter $\hat{\theta}_N^{(k)}$ is obtained as

$$\begin{aligned}\hat{\theta}_N^{(k)} = &\arg\min_{\theta} \\ &\frac{1}{N}\sum_{t=1}^{N}\left[F(q,\theta)\frac{1}{F(q,\hat{\theta}_N^{(k-1)})}y(t) + \right. \\ &\left. -B(q,\theta)\frac{1}{F(q,\hat{\theta}_N^{(k-1)})}u(t)\right]^2.\end{aligned}$$

In view of the notation introduced above, this can be written as

$$\hat{\theta}_N^k = \left[\Phi_F^T(\hat{\theta}_N^{(k-1)})\Phi_F(\hat{\theta}_N^{(k-1)})\right]^{-1}\Phi_F^T(\hat{\theta}_N^{(k-1)})\mathbf{y}_F. \tag{8}$$

As analyzed in Stoica and Söderström (1981) the disadvantage of the Steiglitz-McBride scheme is the fact that even if the real OE-solution $\hat{\theta}_N$ is inserted as $\hat{\theta}_N^{(k-1)}$, the iterative algorithm is not guaranteed to converge to $\hat{\theta}_N$, except when the additive output noise is white and the system is in the model set.

The analysis in the previous section, leading to equation (7), points to an alternative iterative scheme, determined by

$$\hat{\theta}_N^k = \left[\Psi^T(\hat{\theta}_N^{(k-1)})\Phi_F(\hat{\theta}_N^{(k-1)})\right]^{-1}\Psi^T(\hat{\theta}_N^{(k-1)})\mathbf{y}_F. \tag{9}$$

Because of its origin in the previous section, the parameter estimate that results after convergence of this iterative algorithm, necessarily will satisfy $V_N'(\hat{\theta}_N) = 0$, and therefore will always upon convergence arrive at a (local) minimum of the quadratic output error cost function.

In a format that connects to the related expression for the Steiglitz-McBride algorithm, this alternative iterative scheme comes down to

$$\hat{\theta}_N^{(k)} \text{ is the solution to}$$
$$\frac{1}{N}\sum_{t=1}^{N}\left[F(q,\theta)\frac{1}{F(q,\hat{\theta}_N^{(k-1)})}y(t)+ \right. \tag{10}$$
$$\left. -B(q,\theta)\frac{1}{F(q,\hat{\theta}_N^{(k-1)})}u(t)\right]\psi(t,\hat{\theta}_N^{(k-1)}) = 0$$

or equivalently

$$\hat{\theta}_N^{(k)} =$$
$$sol_\theta \frac{1}{N}\sum_{t=1}^{N}\psi(t,\hat{\theta}_N^{(k-1)})[y_F(t) - \varphi_F^T(t,\hat{\theta}_N^{(k-1)})\theta] = 0 \tag{11}$$

It appears that in each step of the iteration the construction of $\hat{\theta}_N^{(k)}$ amounts to applying an instrumental variable (IV) estimator that is applied to filtered input and output data u_F, y_F, and an instrument vector that is determined by $\psi(t,\hat{\theta}_N^{(k-1)})$. Analysis of this instrument vector shows that for an output error model structure (as applied here),

$$\frac{\partial\hat{y}(t|t-1;\theta)}{\partial f_i} = -\frac{B(q,\theta)}{F^2(q,\theta)}u(t-i) \tag{12}$$
$$\frac{\partial\hat{y}(t|t-1;\theta)}{\partial b_i} = \frac{1}{F(q,\theta)}u(t-i), \tag{13}$$

leading to

$$\psi(t,\hat{\theta}_N^{(k-1)}) = \begin{bmatrix} -\tilde{y}_F(t-1) \\ \vdots \\ -\tilde{y}_F(t-n_f) \\ u_F(t-1) \\ \vdots \\ u_F(t-n_b) \end{bmatrix}$$

with $\tilde{y}_F(t) := G(q,\hat{\theta}_N^{(k-1)})u_F(t)$.

The alternative iterative method was mentioned already in the original publication Steiglitz and McBride (1965) as "mode 2 iteration", however it does not seem to have obtained considerable attention in the literature after this publication. In instrumental variable methods, the approach is known under the name of "Simplified Refined IV method" (SRIV), as presented in Young (1976); Young et al. (2008), resulting from the optimality conditions of the maximum likelihood estimator. The particular algorithm is reported to have very good convergence properties, although convergence does not appear to have been proven formally.

Its optimality properties, resulting from the analysis given above, are formulated in the following proposition:

Proposition 1. Consider the data generating system (1) and the output error model structure (3), then the iterative algorithm (11) will upon convergence satisfy $V'(\hat{\theta}_N) = 0$, and therefore will be in a local optimum of the output error cost function.

This result follows from the derivation in Sections 3 and 4. Note that it does not require the assumption that the data generating system is part of the model set. Therefore it also holds in a situation of approximate modelling. As a result this algorithm seems to be much more powerful than the Steiglitz and McBride type of iterative algorithm.

The iterative linear regression algorithms have some resemblance with pseudo-linear regression algorithms that are basically determined by

$$\hat{\theta}_N = sol_\theta \frac{1}{N}\sum_{t=1}^{N}\zeta(t,\theta)[y(t) - \varphi^T(t,\theta)\theta)] = 0. \tag{14}$$

The filtering operation in the iterative regression algorithms (replacing u and y by u_F and y_F) is a distinguishing feature.

5. OPTIMAL IV ESTIMATORS

In this section it will be shown that the iterative algorithm (11) can also be interpreted as implementation of an optimal IV estimator.

Consider the general class of IV estimators:

$$\hat{\theta}_N = sol_\theta \frac{1}{N}\sum_{t=1}^{N} z(t)\cdot L(q)[y(t) - \varphi(t)^T\theta] = 0$$

with $z(t) \in \mathbb{R}^{n_b+n_f}$ a vector of instruments. It has been shown in Söderström and Stoica (1983) that a minimum variance estimator is achieved under the following conditions:

$$z(t) = \frac{1}{H_0(q)F_0(q^{-1})}\tilde{\varphi}(t) \tag{15}$$
$$L(q) = \frac{1}{H_0(q)F_0(q^{-1})} \tag{16}$$

and $\tilde{\varphi}(t)$ is the noise-free part of $\varphi(t)$.

If we evaluate the iterative algorithm presented in the previous section, it appears that it fits into the optimal IV estimator framework through the choices

$$H_0(q) = 1 \tag{17}$$
$$L(q) = \frac{1}{F(q,\hat{\theta}_N^{(k-1)})} \tag{18}$$
$$z(t) = \frac{1}{F(q,\hat{\theta}_N^{(k-1)})}\hat{\varphi}(t) \tag{19}$$

where $\hat{\varphi}(t)$ is an estimate of $\tilde{\varphi}(t)$, determined by

$$\hat{\varphi}(t) = [-\hat{x}(t-1)\cdots-\hat{x}(t-n_f)\ u(t-1)\cdots u(t-n_b)]$$

and $\hat{x}(t) = G(q,\hat{\theta}_N^{(k-1)})u(t)$. In other words: the measured output signals in the regression vector are replaced by outputs that are simulated on the basis of the model estimate from the previous iteration.

The iterative IV algorithm therefore is an implementation of an (approximate) minimal variance estimator for the situation of an Output Error systems ($H_0 = 1$).

6. COMPARISON WITH GRADIENT-BASED OPTIMIZATION

The iterative IV algorithm can be written in a format that clearly exhibits the parameter update mechanism. To this end we denote [4]

$$\hat{\theta}_N^{(k)} = \left[\Psi_{k-1}^T \Phi_{k-1}\right]^{-1} \Psi_{k-1}^T \mathbf{y}_{F,k-1}$$

while

$$\mathbf{y}_{F,k-1} = \Phi_{k-1}\hat{\theta}_N^{(k-1)} + \boldsymbol{\varepsilon}_{k-1}$$

and

$$\boldsymbol{\varepsilon}_{k-1} = [\varepsilon_{oe}(1,\hat{\theta}_N^{(k-1)}) \cdots \varepsilon_{oe}(N,\hat{\theta}_N^{(k-1)})]^T$$

with

$$\varepsilon_{oe}(t,\hat{\theta}_N^{(k-1)}) = y(t) - \frac{B(q^{-1},\hat{\theta}_N^{(k-1)})}{F(q^{-1},\hat{\theta}_N^{(k-1)})}u(t).$$

Combining the equation shows that

$$\hat{\theta}_N^{(k)} = \hat{\theta}_N^{(k-1)} + \left[\Psi_{k-1}^T \Phi_{k-1}\right]^{-1} \Psi_{k-1}^T \boldsymbol{\varepsilon}_{k-1}.$$

Note that a direct gradient-based optimization step as e.g. applied in a Gauss-Newton iteration step, is given by

$$\hat{\theta}_N^{(k)} = \hat{\theta}_N^{(k-1)} + \left[\Psi_{k-1}^T \Psi_{k-1}\right]^{-1} \Psi_{k-1}^T \boldsymbol{\varepsilon}_{F,k-1}$$

where $\Psi_{k-1}^T\Psi_{k-1}$ is an (estimate of) the second derivative $V''(\theta)$ of the cost function in $\hat{\theta}_N^{(k-1)}$, and $\Psi_{k-1}^T\boldsymbol{\varepsilon}_{k-1}$ is the gradient of the cost function $V'(\theta)$ in $\hat{\theta}_N^{(k-1)}$.

The difference between the iterative IV and the Gauss-Newton algorithms is not immediately obvious. It should be noted that the iterative IV algorithm does not rely on an approximate Taylor expansion, and therefore may have improved convergence properties. However results on this issue are referred to future work.

7. RELATED FREQUENCY DOMAIN ALGORITHM

A similar mechanism as used in the time-domain can be used to solve a frequency domain identification problem. In this latter domain the problem is formulated, starting from a given or measured frequency response

$$\{G(e^{i\omega_k})\}_{k=1,\cdots N}$$

that is available in a particular frequency grid denoted by the frequencies ω_k, $k = 1, \cdots N$.

The problem then is to fit a parametric model to this data, by solving the identification problem

$$\hat{\theta} = \arg\min_{\theta} V(\theta) \tag{20}$$

with

$$V(\theta) = \sum_{k=1}^{N} \left| G(e^{i\omega_k}) - \frac{B(e^{i\omega_k},\theta)}{F(e^{i\omega_k},\theta)} \right|^2 \tag{21}$$

and with B, F and θ as defined before.

Whereas this problem is nonconvex if $F(e^{i\omega},\theta)$ is parametrized, an iterative algorithm has been proposed,called the Sanathanan-Koerner algorithm (Sanathanan and Koerner (1963)) that follows the same reasoning as the Steiglitz and McBride in the time domain. It attempts to solve the nonconvex optimization problem by a sequence of convex problems and filter steps by solving

$$\hat{\theta}^{(j)} = \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \left| \frac{1}{F(e^{i\omega_k},\hat{\theta}_N^{(j-1)})} \left[F(e^{i\omega_k},\theta)G(e^{i\omega_k}) - B(e^{i\omega_k},\theta)\right] \right|^2.$$

The properties of this algorithm are similar to the ones for the Steiglitz and McBride algorithm in the time-domain.

If we apply the same philosophy as presented in the earlier sections of this paper, we write

$$\frac{\partial V(\theta)}{\partial \theta} = \sum_{k=1}^{N} -2Re\left\{ M_k^*(\theta)\left[G(e^{i\omega_k}) - \frac{B(e^{i\omega_k},\theta)}{F(e^{i\omega_k},\theta)}\right]\right\}$$

with

$$M_k(\theta) = \frac{\partial \frac{B(e^{i\omega_k},\theta)}{F(e^{i\omega_k},\theta)}}{\partial\theta},$$

and $(\cdot)^*$ denotes complex conjugate transpose.

If we define

$$G_F(e^{i\omega_k}) := \frac{1}{F(e^{i\omega_k},\hat{\theta})}G(e^{i\omega_k})$$

then every $\hat{\theta}$ for which holds that $\frac{\partial V}{\partial\theta} = 0$ will satisfy

$$\sum_{k=1}^{N} Re\left\{ M_k^*(\hat{\theta})\cdot \cdot\left[F(e^{i\omega_k},\hat{\theta})G_F(e^{i\omega_k}) - B(e^{i\omega_k},\hat{\theta})\frac{1}{F(e^{i\omega_k},\hat{\theta})}\right]\right\} = 0.$$

Rewriting this expression in regression format yields

$$\sum_{k=1}^{N} Re\left\{ M_k^*(\hat{\theta})\cdot\left[G_F(e^{i\omega_k}) - \Omega_F(e^{i\omega_k})\hat{\theta}\right]\right\} = 0$$

with

$$\Omega_F(e^{i\omega_k}) = \begin{bmatrix} -e^{-i\omega_k}G_F(e^{i\omega_k}) \\ \vdots \\ -e^{-in_f\omega_k}G_F(e^{i\omega_k}) \\ \frac{-e^{-i\omega_k}}{F(e^{i\omega_k},\hat{\theta})} \\ \vdots \\ \frac{-e^{-in_b\omega_k}}{F(e^{i\omega_k},\hat{\theta})} \end{bmatrix},$$

and this can be reworked into the expression

$$\sum_{k=1}^{N} [Re\{M_k\}\ Im\{M_k\}]\cdot \cdot\left[\begin{bmatrix} Re\{G_F(e^{i\omega_k})\} \\ Im\{G_F(e^{i\omega_k})\}\end{bmatrix} - \begin{bmatrix} Re\{\Omega_F(e^{i\omega_k})\} \\ Im\{\Omega_F(e^{i\omega_k})\}\end{bmatrix}\hat{\theta}\right] = 0.$$

With the notation

$$\overline{M}^T := [Re\{M_1\}\ Im\{M_1\}\ \cdots Re\{M_N\}\ Im\{M_N\}]$$

[4] For brevity we introduce the notation $\Psi_{k-1} = \Psi(\hat{\theta}_N^{(k-1)})$, $\Phi_{k-1} = \Phi_F(\hat{\theta}_N^{(k-1)})$, etc.

and

$$\overline{\Omega} := \begin{bmatrix} Re\{\Omega_F(e^{i\omega_1})\} \\ Im\{\Omega_F(e^{i\omega_1})\} \\ \vdots \\ Re\{\Omega_F(e^{i\omega_N})\} \\ Im\{\Omega_F(e^{i\omega_N})\} \end{bmatrix}, \quad \overline{G}^T := \begin{bmatrix} Re\{G_F(e^{i\omega_1})\} \\ Im\{G_F(e^{i\omega_1})\} \\ \vdots \\ Re\{G_F(e^{i\omega_N})\} \\ Im\{G_F(e^{i\omega_N})\} \end{bmatrix}$$

the solution to the regression equation can be formulated as

$$\hat{\theta} = [\overline{M}^T \overline{\Omega}]^{-1} \overline{M}^T \overline{G}. \tag{22}$$

All matrices on the right hand side of this equation are functions of $\hat{\theta}$, so that the natural iterative identification algorithm becomes

$$\hat{\theta}^{(k)} = [\overline{M}(\hat{\theta}^{k-1})^T \overline{\Omega}(\hat{\theta}^{k-1})]^{-1} \overline{M}(\hat{\theta}^{k-1})^T \overline{G}(\hat{\theta}^{k-1}).$$

Upon convergence of the iterative scheme, the parameter estimate will satisfy (22) and therefore also the first order condition $V'(\hat{\theta}) = 0$. As a result every convergence point of the algorithm will be a local optimum of the quadratic output error cost function (21). And similar to the situation in the time domain, this result does not depend on any assumption on the data generating system, i.e. it also applies to the situation where the system does not belong to the chosen (output error) model class.

8. CONCLUSION

An iterative linear regression algorithm is discussed that upon convergence is able to identify optimal output error models. The iteration involves filtering the data and updating the vector of instruments that is used in the linear regression. In the time domain the method has been introduced before as a particular IV method, but the optimality properties presented here seem to have not been particularly addressed or recognized by the identification community. We have shown that it can be naturally extended to be applicable in the frequency domain also. By its optimality properties the presented algorithm outperforms the classical iterative algorithms of Steiglitz and McBride (time domain) and Sanathanan and Koerner (frequency domain).

REFERENCES

Bayard, D. (1994). High-order multivariable transfer function curve fitting: algorithms, sparse matrix methods and experimental results. *Automatica*, 30(9), 1439–1444.

de Callafon, R., de Roover, D., and Van den Hof, P. (1996). Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. In *Proc. 35th IEEE Conf. Decision and Control*, 2030–2035. Kobe, Japan.

Douma, S. (2006). *From Data to Performance - System Identification Uncertainty and Robust Control Design.* Ph.D. thesis, Delft Univ. Technology, Delft, The Netherlands.

Douma, S. and Van den Hof, P. (2005). An alternative paradigm for probabilistic uncertainty bounding in prediction error identification. In *Proc. 44th IEEE Conf. Decision and Control and European Control Conf., CDC-ECC'05*, 4970–4975. Seville, Spain.

Heuberger, P., Van den Hof, P., and Wahlberg, B. (2005). *Modelling and Identification with Rational Orthogonal Basis Functions.* Springer Verlag.

Ljung, L. (1999). *System Identification: Theory for the User, 2nd Edition.* Prentice-Hall, Englewood Cliffs, NJ, second edition.

Sanathanan, C. and Koerner, J. (1963). Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. Automatic Control*, AC-8, 56–58.

Söderström, T. and Stoica, P. (1983). *Instrumental Variable Methods for System Identification.* Springer Verlag.

Steiglitz, K. and McBride, L. (1965). A technique for the identification of linear systems. *IEEE Trans. Automatic Control*, 10(4), 461–464.

Stoica, P. and Söderström, T. (1981). The steiglitz-mcbride identification algorithm revisited - convergence analysis and accuracy aspects. *IEEE Trans. Automatic Control*, AC-26(3), 712–717.

Young, P. (1976). Some observations on instrumental variable methods of time-series analysis. *Int. J. Control*, 23(5), 593–612.

Young, P., Garnier, H., and Gilson, M. (2008). *Identification of Continuous-time Models from Sampled Data*, chapter Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models, 91–131. Springer, London, UK.

Zhu, Y. (2001). *Multivariable System Identification for Process Control.* Elsevier Science Ltd.

Zhu, Y. (2006). System identification for process control: Recent experience and outlook. In B. Ninness and H. Hjalmarsson (eds.), *Proc. 14th IFAC Symp. System Identification.* Newcastle, NSW, Australia.