On identification methods for direct data-driven controller tuning

Klaske van Heusden[1], Alireza Karimi[1,*] and Torsten Söderström[2]

[1] *Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

[2] *Division of Systems and Control, Department of Information Technology, Uppsala University, P.O. Box 337, SE-75105 Uppsala, Sweden*

SUMMARY

In non-iterative data-driven controller tuning, a set of measured input/output data of the plant is used directly to identify the optimal controller that minimizes some control criterion. This approach allows the design of fixed-order controllers, but leads to an identification problem where the input is affected by noise, and not the output as in standard identification problems. Several solutions that deal with the effect of measurement noise in this specific identification problem have been proposed in literature. The consistency and statistical efficiency of these methods are discussed in this paper and the performance of the different methods is compared. The conclusions offer a guideline on how to solve efficiently the data-driven controller tuning problem .

KEY WORDS: Data-driven controller tuning; Identification for control; Undermodeling

1. INTRODUCTION

The concept of virtual reference controller design was first introduced in [1]. The method allows the design of fixed-order controllers using a single set of measured input/output data. The control objective is formulated as a reference model. Instead of the identification of a plant model, which is then used for controller design, the data is used directly to identify the optimal controller. An advantage is that the order of the controller does not depend on the order of the plant model, i.e. it can be fixed beforehand. However, the approach leads to a non-standard identification problem, where the input is affected by noise in contrast to standard identification problems, where only the output is affected by noise.

Several papers have treated this approach, e.g. [2, 3, 4]. An extension to the original method with an appropriate weighting for fixed-order controllers is presented in [2]. The method is developed for noise-free measurements. For noisy measurements the use of instrumental variables is proposed. In [3] several remarks and extensions to [2] are proposed. The use of a

*Correspondence to: Alireza.Karimi@epfl.ch

Contract/grant sponsor: Swiss National Science Foundation; contract/grant number: 200020-105139

prediction error method (PEM) for the identification of the inverse controller is suggested to deal with measurement noise. In [4] a straightforward tuning scheme is proposed that allows identification of the controller without the need for a virtual reference. A specific choice of extended instrumental variables is proposed that leads to an appropriate frequency weighting and deals with the measurement noise at the same time.

This paper discusses the identification problem encountered in direct data-driven controller tuning in detail. One of the main features of the approach is the possibility to fix the order of the controller and to minimize the 2-norm of an approximate model reference criterion for this class of controllers. In common terms of system identification, this situation leads to undermodeling of the controller. In the case of undermodeling, bias shaping is essential for the achieved control performance [5]. The way the noise enters the controller identification problem complicates this bias shaping, and standard results for prediction error methods are not valid, as shown in Section 3.

The identification methods for the direct controller tuning problem proposed in [2, 3] and [4] are discussed in Sections 4-6. Variance expressions for the different approaches are given if no undermodeling is present and convergence to the optimal solution is analyzed in case of undermodeling. In addition to these three methods, application of the method presented in [6] to the controller tuning problem is presented in Section 7. This method is developed for errors-in-variables problems and uses periodicity of the data. It is shown that this method can also be used in the case of undermodeling. The results are illustrated by a simulation example. The performance of the different approaches is compared for the cases with undermodeling and without undermodeling. The conclusions offer a guideline for the choice of identification method.

2. DIRECT DATA-DRIVEN CONTROLLER TUNING

Direct data-driven controller tuning was initially developed for stable SISO systems [1]. Both stable and unstable SISO systems are considered in [4]. For unstable systems the experiment needs to be performed in closed-loop, for stable systems both open-loop and closed-loop data can be used. This paper considers only stable systems and open-loop experiments.

2.1. Approximate model-reference problem

Consider the stable linear SISO plant $G(q^{-1})$, where q^{-1} denotes the backward shift operator. Specifications for the controlled plant are given as a reference model $M(q^{-1})$. In the following, it is assumed that $M \neq 1$. The control objective is to design the controller $K(\rho)$ such that the closed-loop $\frac{K(\rho)G}{1+K(\rho)G}$ resembles the reference model M, where $\rho \in \mathbb{R}^{n_\rho \times 1}$ is a vector of controller parameters. This can be achieved by minimizing the two-norm of the difference between the reference model and the achieved closed-loop system:

$$J_{mr}(\rho) = \left\| M - \frac{K(\rho)G}{1+K(\rho)G} \right\|_2^2 \tag{1}$$

A discussion on the choice of M can be found in [7]. In the following, M and the structure of $K(\rho)$ are assumed to be appropriate for the system at hand. The model reference criterion (1) is non-convex with respect to ρ. An approximation that is convex for linearly parameterized controllers can be defined using the reference model M, as illustrated next. M can be represented as:

$$M = \frac{K^*G}{1+K^*G}. \tag{2}$$

The backward shift operator is omitted here for convenience and will be omitted in the sequel. K^* is the ideal controller, which is defined indirectly by G and M:

$$K^* = \frac{M}{G(1-M)}. \tag{3}$$

This controller K^* exists since $M \neq 1$. It might be of very high order since it depends on the unknown and possibly high-order plant G. Furthermore, it might not stabilize the plant internally and it might be non-causal.

The ideal sensitivity function is then given by $\frac{1}{1+K^*G} = 1-M$. Using (2), the model reference criterion (1) can be expressed as:

$$J_{mr}(\rho) = \left\| \frac{K^*G - K(\rho)G}{(1+K^*G)(1+K(\rho)G)} \right\|_2^2 \tag{4}$$

Approximation of $\frac{1}{1+K(\rho)G}$ by $1-M$, the ideal sensitivity function, leads to the following approximation of the model reference criterion:

$$J(\rho) = \left\| \frac{K^*G - K(\rho)G}{(1+K^*G)^2} \right\|_2^2 = \left\| (1-M)[M - K(\rho)(1-M)G] \right\|_2^2. \tag{5}$$

The quality of this approximation of $J_{mr}(\rho)$ is discussed in [2]. The controller structure is chosen linear in the parameters ρ,

$$K(\rho) = \beta^T(q^{-1})\rho \tag{6}$$

where

$$\beta(q^{-1}) = [\beta_1(q^{-1}), \cdots, \beta_{n_\rho}(q^{-1})]^T \tag{7}$$

is a vector of size n_ρ of stable linear discrete-time transfer operators (in general an orthogonal basis). Then $J(\rho)$ is a quadratic function of ρ and its global optimizer can be found analytically.

This approximate model-reference criterion in combination with a linearly parameterized controller forms the basis of both the virtual reference feedback tuning (VRFT) [2] and the correlation-based tuning (CbT) approach of [4].

The optimal controller is defined as $K(\rho_0)$ with

$$\rho_0 = \arg\min_{\rho} J(\rho) \tag{8}$$

Two cases are considered in this paper.

C1 The objective can be achieved, i.e. $K^* \in \{K(\rho)\}$. Therefore $K(\rho_0) = K^*$ and $J(\rho_0) = 0$

C2 The objective cannot be achieved, i.e. $K^* \notin \{K(\rho)\}$, $K(\rho_0) \neq K^*$ and $J(\rho_0) > 0$.

C1 is often assumed in system identification. However, since one of the main advantages of data-driven controller design is that the order of the controller can be fixed, this assumption does not necessarily hold and **C2**, the case of undermodeling of the controller, needs to be considered.

2.2. Data-driven approaches

In data-driven controller tuning approaches, a set of measured data from the plant is used directly to minimize a time-domain criterion that approximates (5).

This paper considers only stable systems and open-loop measurement data. Consider a set of input, $r(t)$, and output data, $y(t)$, with data length N from an open-loop experiment. Suppose that the output is generated as:

$$y(t) = G(q^{-1})r(t) + v(t)$$

where $v(t)$ is the measurement noise. To proceed, the following assumptions are considered:

A1 The measurement noise $v(t)$ is uncorrelated with $r(t)$.

A2 The measurement noise can be represented as $v(t) = H_v e(t)$, where $e(t)$ is a zero-mean white noise signal with variance σ^2 and H_v is a stable filter.

A3 The input $r(t) = 0, \forall t \leqslant 0$. $r(t)$ is persistently exciting of order n_ρ and $(1 - M)^2 G$ has no zero on the imaginary axis.

Assumption **A3** is used for the case **C1**. In case **C2**, the control objective is formulated as minimization of $J(\rho)$, the 2-norm of an error function. Because the 2-norm is considered, i.e. the integral of the error function over all frequencies, and $J(\rho) \neq 0, \forall \rho$, the system must be excited at all frequencies. The assumption that the spectrum of $r(t)$, $\Phi_r(\omega) > 0$, $\forall \omega$ is sufficient for the analysis of convergence of the estimate in case **C2**. However, for the ease of notation, the following stronger assumption will be used.

A4 The input $r(t)$ is a zero-mean white noise with unit variance and $(1 - M)^2 G$ has no zero on the imaginary axis.

If $r(t)$ is not white, but $\Phi_r(\omega) > 0$, $\forall \omega$, an additional filter needs to be introduced, which complicates the notation and expressions.

The data-driven controller tuning approaches all use a specific filtering of the signals $r(t)$ and $y(t)$ to define an error signal that is used to minimize (5). The virtual reference feedback tuning approach [2] uses a filter that depends on the inverse of the reference model M. Consequently, M^{-1} needs to be stable in this approach. The extensions proposed in [3] use this same filtering scheme. The scheme proposed in [4] is a straightforward implementation of the approximate model reference criterion, that does not require a virtual reference nor a filter that depends on M^{-1}.

In the following, a generalization of the scheme of [4] is used. The only difference between this generalized scheme and the approach of [4] is that the error is filtered also by the additional filter that was used to construct the instrumental variables of the correlation approach, thereby eliminating all 'hidden' filters. Using this generalized scheme, the same signals can be used in the different identification solutions and a comparison of the different methods can be carried out.

The generalized data-driven controller tuning scheme is shown in Figure 1. The transfer function between $r(t)$ and $\varepsilon_k(t, \rho)$ corresponds to the transfer function in the criterion in (5). The scheme of Figure 1 can be reformulated as shown in Figure 2. This figure clearly shows the nature of the identification problem. The input $y_k(t)$ of the controller to be identified $K(\rho)$

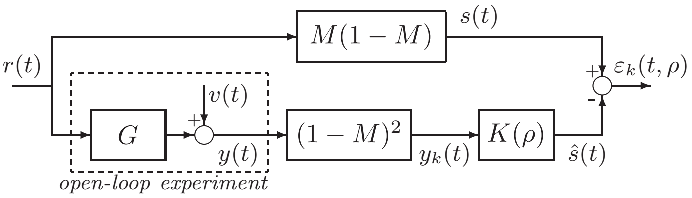

Figure 1. Generalized tuning scheme for model-reference problem using only one experiment

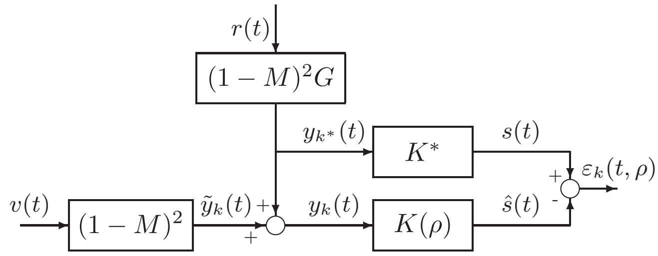

Figure 2. Alternative representation of data-driven controller tuning schemes

is affected by noise $\tilde{y}_k(t)$. The output of the 'true' controller K^* is not affected by noise. Its input, $y_{k^*}(t)$, is also noise-free. The unknown signals in this scheme are $v(t)$, $y_{k^*}(t)$ and the noise signal $\tilde{y}_k(t)$ given by:

$$\tilde{y}_k(t) = (1 - M)^2 v(t) = H_{\tilde{y}} e(t). \tag{9}$$

The known signals are: $r(t)$, $y_k(t) = (1 - M)^2 y(t)$ and $s(t)$ given by:

$$s(t) = (1 - M)^2 G K^* r(t) = (1 - M) M r(t).$$

In contrast to errors-in-variables problems, there is no fundamental identifiability problem since the output $s(t)$ is not affected by noise and the reference signal $r(t)$ is available. However, standard results for PEM approaches are not valid for this specific problem. Furthermore, since one of the main advantages of data-driven controller design is that the order of the controller can be fixed, the case of undermodeling of the controller needs to be considered. The implications of this for the identification problem are illustrated next through application of the PEM approach.

3. IDENTIFICATION OF THE CONTROLLER

Consider the scheme of Figure 2. Using (6), the error can be calculated as

$$\varepsilon_k(t, \rho) = s(t) - \hat{s}(t) = s(t) - K(\rho) y_k(t) = s(t) - \phi^T(t)\rho, \tag{10}$$

where the regression vector $\phi(t)$ is given by:

$$\phi(t) = \beta(q^{-1}) y_k(t) = \beta(q^{-1}) y_{k^*}(t) + \beta(q^{-1}) \tilde{y}_k(t) \triangleq \phi_0(t) + \tilde{\phi}(t). \tag{11}$$

with $\beta(q^{-1})$ defined in (7). In the prediction error framework, the corresponding prediction error is then given by:

$$\begin{aligned}\varepsilon_p(t,\rho,\theta) &= H^{-1}(\theta,\rho)\varepsilon_k(t)\\ &= H^{-1}(\theta,\rho)\big(K^* y_{k^*}(t) - K(\rho)y_{k^*}(t) - K(\rho)\tilde{y}_k(t)\big)\\ &= H^{-1}(\theta,\rho)\big([K^* - K(\rho)]y_{k^*}(t) + [H(\theta,\rho) - K(\rho)H_{\tilde{y}}]e(t)\big) - e(t),\end{aligned} \quad (12)$$

where $H(\theta,\rho)$ is the noise model. Note that the noise filter is given by $K(\rho)H_{\tilde{y}}$, which corresponds to the parameterization of the noise model $H(\theta,\rho)$, i.e. the noise model depends on the controller parameters ρ as well as on the parameters θ.

Direct data-driven controller tuning thus corresponds to an identification problem with a specific parameterization of the noise model, given by $H(\theta,\rho) = K(\rho)H(\theta)$. In the following, the PEM estimate will be analyzed, firstly for case **C1**, where no undermodeling is present. Secondly, the case of fixed-order controllers is treated, where **C1** does not necessarily hold and case **C2** needs to be considered.

*3.1. Case **C1**, $K^* \in \{K(\rho)\}$*

In the standard identification framework, the system to be identified is usually assumed to be contained in the model set, i.e. no undermodeling is present. A well-known result in this case is that, if the noise model and the plant model are independently parameterized, an error in the noise model does not affect consistency of the estimate of the plant model [8]. However, the controller identification considered in this paper leads to a model structure where the controller parameters appear in the noise model. Consequently, the estimate of the controller is consistent only if the noise model is identified correctly.

The PEM estimate with a fixed noise model, i.e. the output error structure where $H(\theta,\rho) = 1$, is now used to illustrate the differences with a standard identification problem. If $H(\theta,\rho) = 1$ and the controller is parameterized as in (6), the PEM criterion $\frac{1}{N}\sum_{t=1}^{N}\varepsilon_p^2(t,\rho)$ is a quadratic function of ρ. The optimizer is given by the least-squares solution:

$$\hat{\rho} = \left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi(t)^T\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi(t)s(t) \quad (13)$$

Under assumption **C1**, it follows from (11) that $s(t) = \phi_0^T(t)\rho_0 = \phi^T(t)\rho_0 - K(\rho_0)\tilde{y}_k(t)$. The estimation error is then given by

$$\hat{\rho} - \rho_0 = \left[\frac{1}{N}\sum_{t=1}^{N}\phi(t)\phi(t)^T\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi(t)K(\rho_0)\tilde{y}_k(t) \quad (14)$$

The regressor $\phi(t)$ is correlated with the noise $\tilde{y}_k(t)$, and therefore the estimate is not consistent.

In contrast to the standard identification problem, the PEM estimate is not consistent here unless the noise model is identified correctly. Similar to the case of closed-loop identification [9], a tailor-made parameterization can be used to find a consistent estimate. If a tailor-made parametrization is used, where $H(\theta,\rho) = K(\rho)H(\theta)$, the estimation error is asymptotically Gaussian distributed [8], i.e. $\sqrt{N}(\hat{\rho} - \rho_0) \overset{dist}{\to} \mathcal{N}(0, P_p)$. The variance of the parameters, which is equal to the Cramér-Rao bound, is given by:

$$P_p = \sigma^2 C_1^{-1} \tag{15}$$

where

$$C_1 = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} [\frac{1}{H^*}\phi_0(t)][\frac{1}{H^*}\phi_0(t)]^T \tag{16}$$

and

$$H^* = K^* H_{\tilde{y}}. \tag{17}$$

We summarize our findings so far:

- The controller tuning problem requires a tailor-made parametrization.
- $(H(\theta)K(\rho))^{-1}$ needs to be stable.
- The identification problem becomes a non-convex optimization problem, also for the linearly parameterized controllers (6).
- If the noise model is not correct, the estimate is not consistent, contrary to the standard Box-Jenkins identification problem.

*3.2. Case **C2**, $K^* \notin \{K(\rho)\}$*

In this case, the criterion $J(\rho) \neq 0$ and the frequency weighting of the error becomes critical for the quality of the controller. This *bias shaping* is well known in the context of iterative identification and control [5].

Asymptotic convergence of the estimate is analyzed next under assumption **A4**. The control objective is defined in (5) and the optimal controller is given by $K(\rho_0)$, with ρ_0 given by (8). The estimate converges asymptotically if

$$\lim_{N\to\infty} \hat{\rho} = \rho_0 = \arg\min_{\rho} J(\rho).$$

The prediction error estimate is given by :

$$\hat{\rho} = \arg\min_{\rho} \frac{1}{N} \sum_{t=1}^{N} \varepsilon_p^2(t,\rho,\theta) = \arg\min_{\rho} J_p(\rho) \tag{18}$$

then

$$\hat{\rho} = \arg\min_{\rho} \frac{1}{N} \sum_{t=1}^{N} \left[H^{-1}(\theta,\rho)\big([K^* - K(\rho)]y_{k^*}(t) + [H(\theta,\rho) - K(\rho)H_{\tilde{y}}]e(t)\big)\right]^2 \tag{19}$$

If no measurement noise is present, the estimate is given by

$$\hat{\rho} = \arg\min_{\rho} \frac{1}{N} \sum_{t=1}^{N} \left[H^{-1}(\theta,\rho)(1-M)\big(M - K(\rho)(1-M)G\big)r(t)\right]^2 \tag{20}$$

The estimate converges asymptotically to the optimal one, only if the noise model is chosen as $H(\theta,\rho) = 1$ (output error structure). In this case, $\lim_{N\to\infty} J_p(\rho) = J(\rho)$ and consequently

$\lim_{N\to\infty} \hat{\rho} = \rho_0$. However, if the measurements are affected by noise, the controller parameters appear in the noise term:

$$\hat{\rho} = \arg\min_{\rho} \frac{1}{N}\sum_{t=1}^{N}\left[H^{-1}(\theta,\rho)\Big((1-M)\big(M-K(\rho)(1-M)G\big)r(t)+\big(H(\theta,\rho)-K(\rho)H_{\tilde{y}}\big)e(t)\Big)\right]^2 \tag{21}$$

Therefore, $H(\theta,\rho)$ should be equal to $K(\rho)H_{\tilde{y}}$ to eliminate the effect of noise and equal to 1 in order to have the appropriate bias shaping. Since these two objectives are in general conflicting $\lim_{N\to\infty} \hat{\rho} \neq \rho_0$.

To summarize: In case **C1**, the PEM gives a consistent estimate, when a tailor-made parametrization is used. This estimate is statistically efficient. However, the PEM cannot be used for bias shaping in case **C2**. Since the possibility to design a low-order controller is one of the main advantages of data-driven controller tuning approaches, case **C2** needs to be considered in practice. A PEM is therefore not an adequate identification method for this specific problem.

4. INSTRUMENTAL VARIABLES

In [2], the use of instrumental variables (IV) is proposed to deal with the measurement noise. The IV solution is given by:

$$\hat{\rho} = \left[\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\phi(t)^T\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\zeta(t)s(t) \tag{22}$$

where $\zeta(t)$ is a vector of length n_ρ that is not correlated with $\tilde{y}_k(t)$.

*4.1. Case **C1**, $K^* \in \{K(\rho)\}$*

Two different choices for the instruments $\zeta(t)$ are discussed in [2] .

Repeated experiment : Perform a second experiment with the same input $r(t)$. The instrumental variable vector is then defined as:

$$\zeta(t) = \beta(q^{-1})y_2(t) = \phi_2(t) \triangleq \phi_0(t) + \tilde{\phi}_2(t).$$

where $\beta(q^{-1})$ is defined in (7). The noise in the second experiment is not correlated to the noise in the first experiment, therefore

$$\lim_{N\to\infty}(\hat{\rho}-\rho_0) = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{t=1}^{N}\phi_2(t)\phi(t)^T\right]^{-1}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\tilde{\phi}_2(t)K(\rho_0)\tilde{y}_k(t) = 0 \tag{23}$$

This estimate is thus consistent.

Identification of the plant : Identify a model of the plant $\hat{G}$ and generate the simulated output $\hat{y}(t) = \hat{G}r(t)$ and define the instruments as $\zeta(t) = \beta\hat{y}(t)$. The estimate is consistent, but the variance depends on the quality of the model [10].

The instruments generated by a second experiment are analyzed in [6]. The estimation error $\sqrt{N}(\hat{\rho}-\rho_0)$ is asymptotically Gaussian distributed and the asymptotic covariance matrix is given by:

$$P_{IV} = \sigma^2 R_0^{-1}(C_2 + C_3)R_0^{-1}, \tag{24}$$

where

$$\begin{aligned} R_0 &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \phi_0(t)\phi_0^T(t) \\ C_2 &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} [H^*\phi_0(t)][H^*\phi_0(t)]^T \\ C_3 &= E[H^*\tilde{\phi}_2(t)][H^*\tilde{\phi}_2(t)]^T \end{aligned} \tag{25}$$

and H^* is defined in (17).

Using a second experiment, $2N$ data points are needed for an estimate with a covariance matrix of $\sim \frac{1}{N}P_{IV}$. Theoretically optimal variance can be achieved by using an optimal instrumental variable [10]. Such optimal instruments depend on the unknown controller parameters. An iterative algorithm can be used to improve the accuracy: in the first iteration a controller is identified with a non-optimal IV, then, in the second iteration, a better IV is constructed based on the controller from the first iteration. This procedure can be continued to improve the accuracy of the estimates. However, the convergence of the iterative algorithm cannot be proven.

4.2. Case ***C2****,* $K^* \notin \{K(\rho)\}$

In contrast to prediction error methods, no identification criterion is minimized in an instrumental variable approach. The instrumental variable estimate is defined directly as (22). However, for the specific choice of IV suggested in [2], where the instrumental variables are generated by a second experiment, an asymptotically corresponding quadratic identification criterion exists. This specific choice of instrumental variables can therefore be used for bias shaping and 2-norm minimization as shown next. Assume that the instruments are generated using a second experiment. The IV solution is then given by:

$$\hat{\rho} = \left[\frac{1}{N}\sum_{t=1}^{N}\phi_2(t)\phi(t)^T\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\phi_2(t)s(t) \tag{26}$$

Since the noise in the second experiment is not correlated with the noise in the first experiment, the estimate converges to

$$\lim_{N\to\infty}\hat{\rho} = \lim_{N\to\infty}\left[\frac{1}{N}\sum_{t=1}^{N}\phi_0(t)\phi_0(t)^T\right]^{-1}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\phi_0(t)s(t)$$

This is the least squares minimum of:

$$J_{IV}(\rho) = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}[s(t)-\phi_0(t)\rho]^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\left[\left((1-M)M-(1-M)^2GK(\rho)\right)r(t)\right]^2 \tag{27}$$

Since $r(t)$ is white, the frequency domain equivalent by Parseval's theorem is given by:

$$J_{IV}(\rho) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|(1-M)M-(1-M)^2GK(\rho)\right|^2 d\omega. \tag{28}$$

where, for the ease of notation, $e^{-j\omega}$ has been omitted from the arguments of M, G and $K(\rho)$. It is clear that $J_{IV}(\rho)$ is equivalent to $J(\rho)$ as defined in (5).

If the instruments are generated using a second experiment, the estimate is consistent. However, the variance of this estimate is relatively large. In order to reduce the variance it is suggested in [2] to generate the instruments by simulating an identified model. In case **C1**, this did not affect the consistency. In case **C2** it does affect convergence. The IV solution in (26) is the minimizer of $J_{IV}(\rho)$ only if the model is identified correctly, i.e. $\hat{G} = G$. If this is not the case, the resulting estimate does not converge to ρ_0.

5. IDENTIFYING THE INVERSE OF THE CONTROLLER

Since the output $s(t)$ is not affected by noise, identification of the inverse of the controller, $K^{-1}(\rho)$, is a standard identification problem. In [3], the use of PEM methods to identify $K^{-1}(\rho)$ is proposed. In [3], the error is constructed using the VRFT scheme and given by:

$$\begin{aligned} \varepsilon_i(t,\rho) &= \frac{L}{M(1-M)}\left(K^{-1}(\rho)s(t) - y_k(t)\right) \\ &= L\left(K^{-1}(\rho) - (K^*)^{-1}\right)r(t) - \frac{L}{M(1-M)}\tilde{y}_k(t) \end{aligned} \tag{29}$$

where L is an appropriate filter. The noise filter is thus given by

$$H_i^* = \frac{L}{M(1-M)}H_{\tilde{y}} \tag{30}$$

5.1. *Case **C1**, $K^* \in \{K(\rho)\}$*

The gradient of $\varepsilon_i(t,\rho)$ is given by:

$$\psi(t) = \frac{d\varepsilon_i(t,\rho)}{d\rho} == -\frac{L\beta(q^{-1})}{K(\rho)^2}r(t) \tag{31}$$

where $\beta(q^{-1})$ is defined in (7). If a model structure is used that identifies a noise model $H(\theta)$ such that $H_i^* \in \{H(\theta)\}$, the covariance matrix is given by [8]:

$$P_i = \sigma^2\left[\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}[\frac{1}{H_i^*}\psi(t)][\frac{1}{H_i^*}\psi(t)]^T\right]^{-1} \tag{32}$$

Replacing H_i^* and $\psi(t)$ by the expressions of (30) and (31) gives:

$$P_i = \sigma^2 C_1^{-1} = P_p, \tag{33}$$

where P_p, defined in (15), corresponds to the Cramér-Rao bound.

Remarks:

- The linear controller structure as defined in (6) now leads to a non-convex optimization problem.
- The inverse of the controller $K^{-1}(\rho)$ needs to be stable.

5.2. Case **C2**, $K^* \notin \{K(\rho)\}$

In this case, *bias shaping* is again essential for the quality of the controller. The following analysis follows the analysis of Section 3. Let the identification criterion for estimation of the inverse of the controller be defined as

$$J_i(\rho) = \frac{1}{N}\sum_{t=1}^{N}[H^{-1}(\theta,\rho)\varepsilon_i(t,\rho)]^2 \tag{34}$$

where $\varepsilon_i(t,\rho)$ is defined as (29) and $H(\theta,\rho)$ is the noise model. If no measurement noise is present, the identification criterion is given by:

$$J_i(\rho) = \frac{1}{N}\sum_{t=1}^{N}\left[H^{-1}(\theta,\rho)L\left(\frac{1}{K(\rho)} - \frac{1}{K^*}\right)r(t)\right]^2 \tag{35}$$

This corresponds to the transfer function in $J(\rho)$ of (5) if $H(\theta,\rho)^{-1}L = K(\rho)M(1-M)$. However, if the measurements were affected by noise, the criterion becomes

$$J_i(\rho) = \frac{1}{N}\sum_{t=1}^{N}\left[K(\rho)M(1-M)\left(\frac{1}{K(\rho)} - \frac{1}{K^*}\right)r(t) - K(\rho)(1-M)^2v(t)\right]^2 \tag{36}$$

In this case the controller parameters appear in the noise term, as was the case for the PEM in section 3 and $\lim_{N\to\infty}\hat{\rho} \neq \rho_0$.

In [3], the use of a different filter $L = M^2/G$ that depends on the unknown plant G is proposed. Clearly L is unknown and only an estimation can be used. The resulting criterion, if G was available, would be

$$J_i(\rho) = \left\|(1-M)M\left[\frac{K^*}{K(\rho)} - 1\right]\right\|_2^2.$$

This criterion does not correspond to $J(\rho)$. In [2] the quality of the approximation $J(\rho)$ is established, the quality of $J_i(\rho)$ remains an open question.

6. CORRELATION APPROACH

The use of a specific choice of extended instrumental variables is proposed in [4]. The method is developed for both periodic and non-periodic data. In the following only the case of non-periodic data is considered. The vector of instrumental variables $\zeta(t)$, correlated with $r(t)$ and uncorrelated with $v(t)$, is defined as:

$$\zeta(t) = [r(t+l_1), \ldots r(t), r(t-1), \ldots, r(t-l_1)]^T \tag{37}$$

where l_1 is a sufficiently large integer. The choice of l_1 is commented on at the end of this section. The correlation criterion $J_{N,l_1}(\rho)$ is defined as

$$J_{N,l_1}(\rho) = \left[\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varepsilon_k(t,\rho)\right]^T\left[\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varepsilon_k(t,\rho)\right] = \sum_{\tau=-l_1}^{\tau=l_1} R_{\varepsilon r}^2(\tau,\rho) \tag{38}$$

where $R_{\varepsilon r}(\tau,\rho)$ is an estimate of the cross-correlation function between the error signal $\varepsilon_k(t,\rho)$ and $r(t)$, defined by:

$$R_{\varepsilon r}(\tau,\rho) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_k(t,\rho)r(t-\tau) \tag{39}$$

It can be shown that, under Assumptions **A1** and **A4** and using Parseval's theorem [4]:

$$\lim_{N,l_1\to\infty,l_1/N\to 0} J_{N,l_1}(\rho) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|(1-M)M-(1-M)^2GK(\rho)\Big|^2 d\omega. \tag{40}$$

The correlation criterion thus converges asymptotically to $J(\rho)$ of (5). The estimate is therefore consistent in case **C1** and converges asymptotically to ρ_0 in case **C2**.

The covariance matrix for the correlation approach is given by [10]:

$$P_c = \sigma^2(Q^TQ)^{-1}Q^TSQ(Q^TQ)^{-1} \tag{41}$$

where

$$Q = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}E\{\zeta(t)\phi^T(t)\} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\phi_0^T(t)$$

$$S = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}[H^*\zeta(t)][H^*\zeta(t)]^T$$

It should be noted that the use of an estimate of the noise filter H^* for constructing the instruments may improve the variance of the parameter estimates for case **C1**. However, it changes the bias distribution of the correlation criterion in (40) for case **C2** and it does not converge to $J(\rho)$ anymore.

The estimate is consistent, but a bias exists when a finite number of data is used for the estimate [4]. The expected value of the criterion is given by:

$$E\{J_{N,l_1}(\rho)\} \approx \tilde{J}_{N,l_1}(\rho) + \frac{\sigma^2(2l_1+1)}{2\pi N}\int_{-\pi}^{\pi}|K(\rho)|^2|H_{\tilde{y}}|^2 d\omega, \tag{42}$$

where $\tilde{J}_{N,l_1}(\rho)$ is the noise-free criterion based on a finite number of data. The design variable l_1 thus leads to a trade-off between the bias due to noise for a finite number of data and the difference $\tilde{J}_{N,l_1}(\rho)-J(\rho)$. l_1 should be chosen sufficiently large for the difference $\tilde{J}_{N,l_1}(\rho)-J(\rho)$ to be small, and sufficiently small to limit the bias (42).

7. PERIODIC ERRORS-IN-VARIABLE APPROACH

The identification problem as shown in Figure 2 can be seen as a specific case of an errors-in-variables problem. Since the output $s(t)$ is noise-free and the reference $r(t)$ is known, there is no fundamental identifiability problem and the identification is a lot simpler than the standard errors-in-variables (EIV) problem. However, techniques developed for EIV problems can be applied to deal with the measurement noise. In particular, the method proposed in [6] is considered, which takes advantage of the periodicity of the reference signal. The method uses an extended IV method.

Assume that the input signal $r(t)$ is periodic with period N_p and contains p periods, i.e. the data length N is equal to pN_p. The deterministic part of $y(t)$ is also periodic (a possible transient in $y(t)$ can be neglected if the number of periods (or data) is large enough). The regression vector $\phi_j(t)$ in period j is defined as :

$$\phi_j(t) = \phi_0(t) + \tilde{\phi}_j(t), \quad t = 1, \dots, N_p, \tag{43}$$

where $1 \leqslant j \leqslant p$ and $\tilde{\phi}_j(t)$ is the noise contribution to the regression vector in period j. $\zeta_j(t)$ denotes the instrumental vector for period j, defined as:

$$\begin{aligned}
\zeta_1(t) &= [\phi_2^T(t) \dots \phi_p^T(t)]^T \\
\zeta_2(t) &= [\phi_3^T(t) \dots \phi_p^T(t)\, \phi_1^T(t)]^T \\
&\vdots \\
\zeta_j(t) &= [\phi_{j+1}^T(t) \dots \phi_p^T(t)\, \phi_1^T(t) \dots \phi_{j-1}^T(t)]^T
\end{aligned} \tag{44}$$

Define the matrices:

$$\begin{aligned}
\bar{\zeta}(t) &= [\zeta_1(t) \dots \zeta_p(t)] \\
\bar{\phi}(t) &= [\phi_1(t) \dots \phi_p(t)]
\end{aligned} \tag{45}$$

and the vector

$$\bar{s}(t) = [s_1(t) \dots s_p(t)]^T \tag{46}$$

where $s_j(t)$ is the output of K^* at time t within period j: $s_j(t) = s(t+(j-1)N_p)$. The solution of the extended IV method proposed in [6] is then given by:

$$\hat{\rho} = (\hat{R}^T \hat{R})^{-1} \hat{R}^T \hat{r} \tag{47}$$

where

$$\begin{aligned}
\hat{R} &= \frac{1}{N} \sum_{t=1}^{N_p} \bar{\zeta}(t) \bar{\phi}^T(t) = \frac{1}{N} \sum_{j=1}^{p} \sum_{t=1}^{N_p} \zeta_j(t) \phi_j^T(t) \\
\hat{r} &= \frac{1}{N} \sum_{t=1}^{N_p} \bar{\zeta}(t) \bar{s}^T(t) = \frac{1}{N} \sum_{j=1}^{p} \sum_{t=1}^{N_p} \zeta_j(t) s_j(t)
\end{aligned} \tag{48}$$

7.1. Case ***C1****,* $K^* \in \{K(\rho)\}$

In [6], it is assumed that the measurement noise within different periods is uncorrelated. If the scheme of Figure 2 is used to generate $s(t)$ and $\phi(t)$, the input is periodic and **A1-A2** are valid, then this is not the case. Even if the measurement noise $v(t)$ is white, i.e. $H_v = 1$, $\tilde{y}_k(t)$ is not white since $H_{\tilde{y}} = (1 - M)^2 H_v$. Consequently, the measurement noise within different periods is correlated. The results of [6] can be extended as follows.

Corollary 1. *Assume that $r(t)$ is periodic and persistently exciting of order n_ρ, that $(1-M)^2 G$ has no zero on the imaginary axis and that **A1-A2** are met. Then, the estimate (47) is consistent and the asymptotic covariance matrix of the estimation error $\sqrt{N}(\hat{\rho} - \rho_0)$ for (47) is given by:*

$$P_{EIV} = \sigma^2 R_0^{-1} \left(C_2 + \frac{C_3}{p-1} \right) R_0^{-1}, \tag{49}$$

where C_2 is defined as in (25) and

$$C_3 = E\left\{[H^*\tilde{\phi}_j(t)][H^*\tilde{\phi}_j(t)]^T\right\} \tag{50}$$

Proof 1. *The main idea of the proof is that, as $N_p \to \infty$, the noise within different periods is uncorrelated. The full proof is given in the Appendix.*

Note that the definition of C_3 corresponds to the definition in (25), with $\tilde{\phi}_2$ replaced by $\tilde{\phi}_j$ and, if the characteristics of the noise are the same, these matrices are equivalent. It is shown in [6] that the achieved variance P_{EIV} is optimal in the class of all possible extended IV methods. As the number of periods $p \to \infty$, $P_{EIV} \to \sigma^2 R_0^{-1} C_2 R_0^{-1}$, which is equal to the variance for the output error structure in a standard identification problem, where the output is affected with noise [8]. This variance is the optimal variance that can be achieved when no noise model is identified, $P_{opt} = \sigma^2 R_0^{-1} C_2 R_0^{-1}$. As the number of periods $p \to \infty$, the variance thus converges to this optimal variance, $P_{EIV} \to P_{opt}$.

*7.2. Case **C2**, $K^* \notin \{K(\rho)\}$*

The estimate in (47) is the optimum of

$$J_{EIV}(\rho) = \|\hat{r} - \hat{R}\rho\|_2^2 \tag{51}$$

Define

$$r_0 = \frac{1}{N_p}\sum_{t=1}^{N_p} \phi_0(t)s(t) \tag{52}$$

and note that, for periodic data, R_0 is equivalent to

$$R_0 = \frac{1}{N_p}\sum_{t=1}^{N_p} \phi_0(t)\phi_0^T(t) \tag{53}$$

The matrix $\hat{R}$ converges to

$$\lim_{N\to\infty} \hat{R} = [R_0 \ldots R_0]^T \tag{54}$$

Equivalently

$$\lim_{N\to\infty} \hat{r} = [r_0 \ldots r_0]^T \tag{55}$$

and

$$\lim_{N\to\infty} J_{EIV}(\rho) = \left\| \begin{bmatrix} r_0 \\ \vdots \\ r_0 \end{bmatrix} - \begin{bmatrix} R_0 \\ \vdots \\ R_0 \end{bmatrix} \rho \right\|_2^2 = (p-1)\,\|r_0 - R_0\rho\|_2^2 \tag{56}$$

Since R_0 is square and has full rank, the optimum of this criterion is given by

$$\lim_{N\to\infty} \hat{\rho} = R_0^{-1} r_0 \tag{57}$$

This is the solution of the following least-squares criterion

$$R_0^{-1} r_0 = \arg\min_{\rho} \frac{1}{N_p}\sum_{t=1}^{N_p}(s(t) - \phi_0(t)\rho)^2.$$

If the periodic excitation $r(t)$ is chosen such that $\Phi_r(\omega_k) = 1$ for $\omega_k = 2\pi k/N_p$ and $k = 0, \ldots, N_p - 1$, it follows from (27) and (28), that, as $N_p \to \infty$, the minimizer of $J_{EIV}(\rho)$ converges to ρ_0 in case **C2**. Note that such a signal can be generated as a multi-sine, or as a PRBS signal with an offset.

8. COMPARISON

*8.1. Case **C1**, $K^* \in \{K(\rho)\}$*

Under assumption **C1**, the following can be concluded:

- The Cramér-Rao bound can be achieved when a PEM with a tailor-made parameterization is used. The noise-model needs to be identified correctly for consistency, in contrast to the standard identification problem.
- The Cramér-Rao bound can also be achieved when identifying the inverse of the controller. In this case the noise-model does not affect consistency.
- The Cramér-Rao bound can also be achieved by using optimal instrumental variables.

These methods lead to a non-convex optimization problem (also for a linearly parameterized controller). Convergence to the global optimum cannot in general be guaranteed. Furthermore, the inverse of the controller needs to be stable.

The correlation approach and the errors-in-variables approach lead to a convex optimization when the controller is parameterized as in (6). No noise model is identified.

- The errors-in-variables approach is optimal, as the number of periods $p \to \infty$, i.e. $P_{EIV} \to P_{opt}$.
- The variance expression for the correlation method is difficult to analyze. Under some specific hypotheses it can be shown that the variance of an extended IV tends to the optimal variance [10], but for the general identification problem this is not the case. The design parameter l_1 affects the bias with respect to noise for a finite number of data.

To conclude, under assumption **C1**, the Cramér-Rao bound can be achieved. Since the noise model does not affect consistency when the inverse of the controller is identified, identifying the inverse of the controller should be prefered over the use of a PEM approach. If the inverse of the controller is unstable, the best variance achievable, P_{opt}, is achieved using the errors-in-variables approach, when $p \to \infty$.

*8.2. Case **C2**, $K^* \notin \{K(\rho)\}$*

Assumption **C1** is not compatible with one of the main motivations for direct controller tuning: tuning of controllers of limited order. If the order of the controller is fixed beforehand, and the controller minimizing a 2-norm is sought, **C1** is per definition violated and case **C2** needs to be considered. It is shown in Section 3 that in this case, the estimate by PEM does not converge asymptotically to ρ_0 . In Section 5 it is shown that identification of the inverse of the controller also does not converge to the optimal solution. It is thus necessary to resort to the statistically less efficient methods using (extended) instrumental variables.

Clearly,

$$P_{IV} = \sigma^2 R_0^{-1}(C_2 + C_3)R_0^{-1} > \sigma^2 R_0^{-1}\left(C_2 + \frac{C_3}{p-1}\right)R_0^{-1} = P_{EIV}. \quad (58)$$

The instrumental variable approach as proposed by [2] will therefore not be considered in the following, neither in the simulation example.

In case **C2**, the frequency weighting is essential. Since the identification of a noise model affects the frequency weighting, no noise model can be used when bias shaping is required. The optimal variance that can be achieved in this case is given by P_{opt}, see Section 7.1. The errors-in-variables method leads to this optimal variance P_{opt}, as $p \to \infty$. If the reference signal $r(t)$ can be chosen, it is therefore advisable to use a periodic signal with many periods and use the EIV method. If the reference signal cannot be chosen arbitrarily and only few periods or no periodic data is available, the correlation approach should be used.

To conclude, in case **C2**, the price to pay for convergence is a loss in statistical efficiency, since no noise model can be used to improve the estimate. The achieved control objective $J(\hat{\rho})$ depends on both the variance and the bias error and this mean-square-error of the criterion depends strongly on the nature of the problem, i.e. the reference signal $r(t)$, the noise spectrum and the distance between K^* and $K(\rho)$.

9. SIMULATION EXAMPLE

The different methods discussed are tested in simulation on the flexible transmission system proposed as benchmark in [11]. This example was used in [2, 3] and [4] to illustrate the direct data-driven controller tuning approach.

The plant is given by the discrete-time model $G(q^{-1})$

$$G(q^{-1}) = \frac{0.283q^{-3} + 0.507q^{-4}}{A(q^{-1})}$$

with

$$A(q^{-1}) = 1 - 1.418q^{-1} + 1.589q^{-2} - 1.316q^{-3} + 0.886q^{-4}$$

The controller structure is given as

$$K(\rho) = \frac{\rho_1 + \rho_2 q^{-1} + \rho_3 q^{-2} + \rho_4 q^{-3} + \rho_5 q^{-4} + \rho_6 q^{-5}}{1 - q^{-1}}$$

PRBS signals with unity amplitude are used as input, $r(t)$, to the system. The output of the plant is disturbed by zero-mean white noise, such that the signal-to-noise ratio of the open-loop experiment is about 10 in terms of variance. The first periods, i.e. from zero initial state, are used in the PEM method and when identifying the inverse of the controller. For the correlation approach and the errors-in-variables method, a periodic signal of the same length is used. Since $r(t)$ is a PRBS signal, the extended instruments of (37) can be taken as, $\zeta(t) = [r(t), r(t-1), \ldots, r(t-l_1)]^T$. In the following, $l_1 = 25$, for which $\tilde{J}_{N,l_1}(\rho)$ is a good approximation of $J(\rho)$.

Results are given for different period lengths N_p and an increasing number of periods p. A Monte-Carlo simulation with 100 experiments is performed for each input signal $r(t)$, using a different noise realization for each experiment. The noise realizations are the same for all methods.

9.1. *Case* **C1**, $K^* \in \{K(\rho)\}$

The reference model is defined as

$$M_1(q^{-1}) = \frac{K(\rho_0)G}{1+K(\rho_0)G} \tag{59}$$

with

$$\rho_0 = [0.2045 \quad -0.2715 \quad 0.2931 \quad -0.2396 \quad 0.1643 \quad 0.0084]^T \tag{60}$$

The optimal controller $K(\rho_0) \in \{K(\rho)\}$ and the objective can be achieved. The PEM approach with a tailor-made parameterization has not been implemented. Instead, the Box-Jenkins structure is used, which should theoretically be consistent if the order of the noise model is large. The inverse of the controller is identified using the Box-Jenkins structure (Inv).

The results are given in Table I. As expected, estimation of the inverse is efficient. The error for the correlation approach (CbT) and the periodic errors-in-variables approach (EIV) is about 2 times bigger than that of the identification of the inverse of the controller. However, the PEM does not perform as expected. The estimate of the noise model is not exact, therefore the estimate of the controller is biased. Although this method is consistent, for a finite number of data it is not efficient. A tailor-made parameterization would probably perform better. The EIV approach is more efficient than the correlation approach.

Table I. Mean values for the achieved performance $J(\rho)$, for M_1 and different period lengths N_p.

	$N_p = 63$, $p = 16$	$N_p = 127$, $p = 8$	$N_p = 127$, $p = 12$
PEM	0.01240	0.00730	0.00729
Inv	0.00466	0.00324	0.00227
CbT	0.00784	0.00643	0.00545
EIV	0.00676	0.00558	0.00472

A closer look at the CbT and the EIV solutions shows that the main difference between the two estimates is the bias (42), see Table II. The standard deviation of the EIV estimate is slightly bigger than that for CbT, however, the mean error is larger for CbT due to the bias.

Table II. Estimated parameters, the error is given by mean$(\hat{\rho} - \rho_0)$ for CbT and EIV approaches

	CbT approach		EIV approach	
	mean error	std	mean error	std
ρ_1	−0.00223	0.00366	−0.00065	0.00443
ρ_2	0.00640	0.00842	0.00204	0.01105
ρ_3	−0.00909	0.01103	−0.00290	0.01490
ρ_4	0.00901	0.01129	0.00285	0.01496
ρ_5	−0.00638	0.00902	−0.00200	0.01115
ρ_6	0.00234	0.00406	0.00080	0.00458

9.2. Case ***C2****,* $K^* \notin \{K(\rho)\}$

The control objective is defined by the closed-loop reference model

$$M_2(q^{-1}) = \frac{q^{-3}(1-\alpha)^2}{(1-\alpha q^{-1})^2}$$

with $\alpha = 0.606$. In this case $K^* \notin \{K(\rho)\}$ and the objective cannot be achieved. However, this problem can be considered well-defined. Even though the optimal controller cannot be found, the error $M - K(\rho)G(1-M)$ can be made small and the optimal fixed-order controller $K(\rho_0)$ achieves good closed-loop performance. The results are given in Table III.

The results for the estimation of the inverse are not acceptable. These results show that the bias shaping is essential in case **C2**, even if the distance between $K(\rho)$ and K^* is relatively small. The PEM does not converge either, and the performance of the converging estimates CbT and EIV is better. CbT and EIV give again similar performance. As in case **C1**, the error for CbT is mainly due to the bias, whereas the error for EIV is mainly due to variance.

Table III. Mean values for the achieved performance $J(\rho)$, for M_2.

	$N_p = 63$, $p = 16$	$N_p = 127$, $p = 8$	$N_p = 127$, $p = 12$
PEM	0.1587	0.0791	0.0791
Inv	30.9602	32.9735	33.0103
CbT	0.0754	0.0744	0.0743
EIV	0.0748	0.0743	0.0742

10. CONCLUSION

This paper compares different identification methods that have been proposed for non-iterative data-driven controller tuning. Two distinctive cases are considered. In the first case it is assumed that the objective can be achieved, i.e. there is no undermodeling of the controller. In this case, the Cramér-Rao bound can be attained when the inverse of the controller is identified. However, non-convexity of the optimization might lead to numerical problems and the inverse of the controller is required to be stable. The correlation approach (CbT) and periodic errors-in-variables method (EIV) are convex for a linearly parameterized controller. However, these methods are statistically less efficient since no noise model is estimated.

In practice, the control objective cannot be achieved when a fixed-order controller is designed and undermodeling of the controller needs to be taken into account. In this case, bias shaping is essential. One has to resort to instrumental variable approaches to guarantee convergence of the estimate. It is shown that the performance achieved by CbT and EIV is better than the performance of the other solutions treated in this paper. The choice of the optimal method depends on the nature of the available data, i.e. periodicity of the signals and the number of periods available.

REFERENCES

1. G. O. Guardabassi and S. M. Savaresi, "Virtual reference direct design method: an off-line approach to data-based control system design," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 954–959, 2000.
2. M. C. Campi, A. Lecchini, and S. M. Savaresi, "Virtual reference feedback tuning: A direct method for the design of feedback controllers," *Automatica*, vol. 38, pp. 1337–1346, 2002.
3. A. Sala and A. Esparza, "Extensions to "virtual reference feedback tuning: A direct method for the design of feedback controllers"," *Automatica*, vol. 41, no. 8, pp. 1473–1476, 2005.
4. A. Karimi, K. van Heusden, and D. Bonvin, "Non-iterative data-driven controller tuning using the correlation approach," in *European Control Conference*, Kos, Greece, 2007, pp. 5189–5195.
5. M. Gevers, "Identification for control: from the early achievements to the revival of experiment design," in *44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, December 2005.
6. T. Söderström and M. Hong, "Identification of dynamic errors-in-variables systems with periodic data," in *16th IFAC World Congress*, Prague, Czech Republic, 2005.
7. A. Dehghani, A. Lanzon, and B. D. Anderson, "H_∞ design to generalize internal model control," *Automatica*, vol. 42, no. 11, pp. 1959 – 1968, 2006.
8. L. Ljung, *System Identification - Theory for the User*, 2nd ed. NJ, USA: Prentice Hall, 1999.
9. U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215 – 1241, 1999.
10. T. Söderström and P. Stoica, *System Identification.* U.K.: Prentice-Hall, 1989.
11. I. D. Landau, D. Rey, A. Karimi, A. Voda, and A. Franco, "A flexible transmission system as a benchmark for robust digital control," *European Journal of Control*, vol. 1, no. 2, pp. 77–96, 1995.

APPENDIX

I. PROOF OF COROLLARY 1

The estimate of equation (47) is consistent if $\lim_{N_p\to\infty}\hat{\rho}=\rho_0$, where $K(\rho_0)=K^*$. Using (10) and (11), $s(t)$ can be written as

$$s(t)=\phi^T(t)\rho_0-K(\rho_0)\tilde{y}_k(t),$$

where $\tilde{y}_k(t)$ is defined in (9). Define

$$\hat{q}=-\frac{1}{N_p p}\sum_{j=1}^{p}\sum_{t=1}^{N_p}\zeta_j(t)K(\rho_0)\tilde{y}_{k_j}(t),$$

where $\zeta_j(t)$ is defined in (44). The error $\hat{\rho}-\rho_0$ is then given by:

$$\hat{\rho}-\rho_0=(\hat{R}^T\hat{R})^{-1}\hat{R}^T\hat{q}$$

Consistency is guaranteed if $\lim_{N_p\to\infty}\hat{q}=0$ and $\lim_{N_p\to\infty}\hat{R}=e_{p-1}\otimes R_0$, where $e_{p-1}=(1\dots1)^T$ has dimension $(p-1)\times 1$ and $\otimes$ denotes Kronecker product [6]. Convergence of

$\lim_{N_p\to\infty}\hat{q}$ is shown next. Substituting $\zeta_j(t)$ by (44) gives:

$$\begin{aligned}\lim_{N_p\to\infty}\hat{q} &= \lim_{N_p\to\infty}\left(-\frac{1}{N_p p}\sum_{j=1}^{p}\sum_{t=1}^{N_p}[\phi_{j+1}^T(t)\ldots\phi_p^T(t)\,\phi_1^T(t)\ldots\phi_{j-1}^T(t)]^T K(\rho_0)\tilde{y}_{k_j}(t)\right)\\ &= \lim_{N_p\to\infty}\left(-\frac{1}{N_p p}\sum_{j=1}^{p}\sum_{t=1}^{N_p}[\tilde{\phi}_{j+1}^T(t)\ldots\tilde{\phi}_p^T(t)\,\tilde{\phi}_1^T(t)\ldots\tilde{\phi}_{j-1}^T(t)]^T K(\rho_0)\tilde{y}_{k_j}(t)\right)\\ &= \lim_{N_p\to\infty}\left(-\frac{1}{N_p p}\sum_{j=1}^{p}\sum_{t=1}^{N_p}[(\beta\tilde{y}_{k(j+1)}(t))^T\ldots(\beta\tilde{y}_{kp}(t))^T\,(\beta\tilde{y}_1(t))^T\ldots(\beta\tilde{y}_{k(j-1)}(t))^T]^T K(\rho_0)\tilde{y}_{k_j}(t)\right).\end{aligned} \tag{61}$$

According to (61), $\hat{q}$ consists of cross-correlations between $\beta\tilde{y}_k(t)K(\rho_0)\tilde{y}_k(t-\tau)$, with $\tau = N_p(j-n), n = [1\ldots j-1, j+1\ldots p]$. Note that, if $j=1$, $n=[j+1\ldots p]=[2\ldots p]$ and if $j=p$, $n=[1\ldots p-1]$. The minimal value for τ that appears in $\hat{q}$ is thus $\tau = N_p$. Under assumption **A1-A2**,

$$\lim_{N_p\to\infty}\frac{1}{N_p}\sum_{t=1}^{N_p}\beta(q^{-1})\tilde{y}_k(t)K(q^{-1},\rho_0)\tilde{y}_k(t-\tau) = \beta(q^{-1})H_{\tilde{y}}(q^{-1})K(q,\rho_0)H_{\tilde{y}}(q)\mathrm{X}_e(\tau), \tag{62}$$

where

$$\mathrm{X}_e(\tau) = \lim_{N_p\to\infty}\frac{1}{N_p}\sum_{t=1}^{N_p}e(t)e(t-\tau) = E\{e(t)e(t-\tau)\},$$

the autocorrelation of $e(t)$. Note that, throughout the paper, the backward shift operator q^{-1} is used and omitted for convenience. In (62), the shift operator is mentioned explicitly, because both q^{-1} and q appear. Under assumption **A2**, $\mathrm{X}_e(\tau) = \sigma^2$, for $\tau = 0$ and $\mathrm{X}_e(\tau) = 0$, for $\tau \neq 0$. By definition, $K(q^{-1},\rho_0)$ and $K(q,\rho_0)$ are stable. $H_{\tilde{y}}(q)$, $H_{\tilde{y}}(q^{-1})$ and $\beta(q^{-1})$ are also stable, therefore

$$\lim_{N_p\to\infty}\frac{1}{N_p}\sum_{t=1}^{N_p}\beta\tilde{y}_k(t)K(\rho_0)\tilde{y}_k^T(t-\tau)\to 0, |\tau|\to\infty.$$

Consequently, $\lim_{N_p\to\infty}\hat{q}\to 0$. Convergence of $\lim_{N_p\to\infty}\hat{R}$ can be shown using similar reasoning and is omitted here. Validity of (49) follows from applying the proof above to the results of [6], i.e. asymptotically the results of [6] hold.