C, 8 8

# FREQUENCY DOMAIN EXPRESSIONS OF THE ACCURACY OF A MODEL-FREE CONTROL DESIGN SCHEME

Håkan Hjalmarsson* Michel Gevers**

* *Dept. of Signals, Sensors and Systems, The Royal Institute of Technology, S-100 44 Stockholm, Sweden*
** *CESAME, Bâtiment Euler, Louvain University, B-1348 Louvain-la-Neuve, Belgium*

**Abstract:** Recently a model-free control design scheme, Iterative Feedback Tuning (IFT), has been proposed which is based on a Gauss-Newton search of the parameter space where the search direction is computed from *experimental* data. In this contribution we derive explicit frequency domain expressions for the accuracy of the algorithm when a fixed step-size is used when the design objective is minimum variance control. It is shown that the controller mismatch due to finite data effects depends only on the noise spectrum. The performance of the scheme is also compared with the best known model-based approach: system identification, using optimal input design, followed by controller design based on the identified model. The analysis is asymptotic in the orders of the plant, noise dynamics and controller.

**Keywords:** Minimum-variance control, Frequency domain, High-order systems.

## 1. INTRODUCTION

Recently, iterative identification and control design schemes have received considerable attention, see e.g. Zang *et al.* (1995), Schrama (1992) and Lee *et al.* (1993). These schemes iteratively perform plant model identification and model-based controller update in the closed loop. Iterative Feedback Tuning (IFT), Hjalmarsson *et al.* (1994), is a continuation of these ideas where it is shown that, for linear plants and certain control criteria, *e.g.* LQG, it is possible to carry out the optimization of the controller parameters using measurements from the plant collected during (essentially) normal operating conditions. No modeling of the plant and the disturbance are required. It is thus an alternative to identification and control.

IFT is known to have the stationary points of the control criterion as its only possible points of convergence. A typical scenario for this scheme is thus that, eventually, the neighbourhood of a local minimum of the control criterion will be reached. It is then of interest to know how close to the (locally) optimal controller one will get for a given step-size of the scheme. This is the theme of this paper. It is possible to derive expressions for the controller accuracy for an arbitrary plant but these expressions depend on quantities which must be computed on a case by case basis (such as the optimal controller) and furthermore these expressions do not easily lend themselves to an interpretation. However, by assuming the orders of the controller, the plant and the noise dynamics to be high, it is possible to derive frequency domain expressions for the accuracy of the controller and the loss of performance due to the imprecise estimate. By furthermore considering a simple control problem

[1] M. Gevers acknowledges the support of the Inter University Poles of Attraction of the Belgian State, Prime Minister's Office.

such as minimum variance and assuming that the controller has the same order as the plant and the noise dynamics it is possible to obtain frequency domain expressions for the accuracy of the controller and the performance degradation which are explicit functions of the true system and noise dynamics.

In Section 2 minimum variance control is briefly discussed. The IFT method is outlined in Section 3. The accuracy of IFT is then discussed in Section 4. Section 5 contains a comparison with system identification where optimal input design is used.

## 2. MINIMUM VARIANCE CONTROL

Let the input $u(t)$ and output $y(t)$ of the true time-discrete system be related by

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

where $G_0(q)$ is a linear time-invariant transfer function [2], where $H_0(q)$ is a monic, stable and inversely stable transfer function which is a spectral factor of the process noise $v$ and where $\{e(t)\}$ is zero mean white noise with variance $\sigma^2$.

We will use the following one degree of freedom controller:

$$u(t) = C(q,\rho)(y(t) - r(t)) \tag{2}$$

where $C(q,\rho)$ is a linear time-invariant transfer function of some fixed degree which is parameterized by a parameter vector $\rho$, representing the controller parameters that are available to the user. The reference signal $r(t)$ should be zero under normal operation but has been included since it is used in IFT, see below. To ease the notation somewhat we will from now on omit the time argument of the signals unless needed. In addition, whenever signals are obtained from the closed loop system with the controller $C(\rho)$ operating, we will indicate this by using $\rho$ as argument; thus, $y(\rho)$ will denote the output of the system (1) in feedback with the controller (2).

The output from the closed loop with $r \equiv 0$ is given by

$$y(\rho) = \frac{1}{1 + G_0C(\rho)}v. \tag{3}$$

Consider now the problem of designing a controller that minimizes the output variance (minimum variance control). For this purpose we introduce the control criterion

$$J(\rho) = \mathrm{E}[y^2(\rho)] \tag{4}$$

where E denotes expectation over $e$.

It is evident from (3) that $J(\rho)$ depends in a fairly complicated way on $\rho$. Therefore, there is in general no explicit solution to the minimization problem

$$\rho_{opt} = \arg\min_{\rho} J(\rho) \tag{5}$$

However, when the structure of the controller is unconstrained the solution to this problem can be obtained by e.g. spectral factorization, see e.g. Åström and Wittenmark (1984). A special case is when the true system $G_0$ contains exactly one pure time-delay and is minimum phase. Then the minimum variance controller is given by

$$C_{opt} = \frac{H_0 - 1}{G_0} \tag{6}$$

However, if the structure of the controller is restricted (and not equal to the structure of the overall optimal one) one has to resort to numerical minimization. Such a procedure can be described in the following way:

Find a solution for $\rho$ to the equation

$$0 = J'(\rho) = \mathrm{E}[y(\rho)y'(\rho)]. \tag{7}$$

where $'$ denotes differentiation wrt $\rho$.

A numerical approach to this problem is to take repeated steps in a descent direction

$$\rho_{i+1} = \rho_i - \gamma_i R_i^{-1} J'(\rho_i). \tag{8}$$

Here $R_i$ is some appropriate positive definite matrix, typically an estimate of the Hessian of $J$, such as a Gauss-Newton approximation of this Hessian.

## 3. MODEL-FREE OPTIMAL TUNING

The crucial quantity that is needed in the numerical optimization scheme outlined in Section 2 is

$$J'(\rho_i) = \mathrm{E}[y(\rho_i)y'(\rho_i)]$$

To compute this quantity requires that the true system $G_0$, the noise dynamics $H_0$ and the noise variance $\sigma^2$, are known. The problem we are interested in here is when we are given access to a plant (which is unknown) for experimentation and our task is to design a minimum variance controller for a controller of a certain structure, such as a PID controller. Hence, the scheme (8) is not feasible.

In Iterative Feedback Tuning $y(\rho_i)$ is computed exactly and $y'(\rho_i)$ is computed approximately using experimental data from (essentially) normal operating conditions only, with the following procedure:

In the $i$th iteration, where the controller $C(\rho_i)$ is available, perform first an experiment of length $N$ with controller $C(\rho_i)$ in the loop and reference signal $r \equiv 0$. This gives

$$y_1(t,\rho_i) = \frac{H_0}{1 + G_0C(\rho_i)}e_1(t), \quad t = 1, 2, \ldots, N$$

[2] $q$ denotes the shift operator.

as output where $\{e_1(t)\}_{t=1}^N$ is a realization of the white noise process $e$ in (1).

Next, perform a second experiment with controller $C(\rho_i)$ in the loop and reference signal $r(t) = F(q)y_1(t,\rho_i)$, $t = 1,2,\ldots,N$ where $F(q)$ is an arbitrary stable minimum phase prefilter. This gives the output

$$y_2(t,\rho_i) = \frac{G_0C(\rho_i)F}{1+G_0C(\rho_i)}y_1(t,\rho_i) + \frac{H_0}{1+G_0C(\rho_i)}e_2(t)$$

where $\{e_2(t)\}_{t=1}^N$ is a realization of the white noise process $e$ in (1) which is independent of $\{e_1(t)\}_{t=1}^N$. Now compute

$$\hat{y}'(t,\rho_i) = -\frac{C'(\rho_i)}{C(\rho_i)}F^{-1}y_2(t,\rho_i)$$

It can easily be shown that

$$\hat{y}'(t,\rho_i) = y_1'(t,\rho_i) - \frac{C'(\rho_i)}{C(\rho_i)}\frac{H_0F^{-1}}{1+G_0C(\rho_i)}e_2(t)$$

Notice that the disturbance $e_2$ in the second experiment perturbs the gradient estimate. In the next section the consequences of this are analyzed.

Next compute

$$\hat{J}_N'(\rho_i) = \frac{1}{N}\sum_{t=1}^{N}\hat{y}'(t,\rho_i)y_1(t,\rho_i)$$

This is an unbiased estimate of $J'(\rho_i)$, i.e. $\mathrm{E}[\hat{J}_N'(\rho_i)] = J'(\rho_i)$.

Finally compute

$$\rho_{i+1} = \rho_i - \gamma_i R_i^{-1}\hat{J}_N'(\rho_i). \qquad (9)$$

and continue (if desired) with the next iteration.

Thus, iteration $i$ consists of two closed-loop experiments with the controller $C(\rho_i)$ in the loop. The first experiment uses no reference signal while the other experiment uses the output of the first experiment as reference signal.

Due to the unbiasedness of the gradient estimate, this scheme is of stochastic approximation type, see Robbins and Monro (1951) and Benveniste *et al.* (1987), and will typically end up close to a stationary point of the criterion $J(\rho)$ if the step-size $\gamma_i$ is choosen appropriately. Furthermore, by letting the step-size tend to zero sufficiently fast, the estimate will converge to a stationary point. For more details on the method we refer the reader to Hjalmarsson *et al.* (1994).

*Modification of search direction*

There are many possible choices for the update direction $R_i$ in (9). The identity matrix gives the negative gradient direction. An interesting choice is

$$R_i = \frac{1}{N}\sum_{t=1}^{N}\hat{y}'(t,\rho_i)\,[\hat{y}'(t,\rho_i)]^T \qquad (10)$$

for which the signals are available from the experiments described above. This will give a biased (due to the disturbance in the second experiment) approximation of the Gauss-Newton direction. The algorithm with this choice of update direction will be analyzed in the next section.

## 4. ASYMPTOTIC ACCURACY

In practice one will only perform a finite number of iterations and it is then of interest to analyze the accuracy with which the method is able to find the optimal parameter vector. We shall be particularly interested in the following quantities:

i) The accuracy of the controller parameters
$$\mathrm{E}[(\rho-\rho_{opt})(\rho-\rho_{opt})^T]$$
where $\rho$ denotes the final controller parameter.

ii) The accuracy of the controller in the frequency domain
$$\mathrm{E}[|C(e^{i\omega},\rho) - C(e^{i\omega},\rho_{opt})|^2]$$

iii) The performance degradation
$$J_V(\rho) = \mathrm{E}[(y(t,\rho) - y(t,\rho_{opt}))^2]$$

In order to make the analysis tractable we shall make the following assumptions:

S1) The linear time-invariant system $G_0$ is of finite order, $n$ say, and contains exactly one time delay and is minimum-phase.

S2) The noise dynamics $H_0$ is monic, exponentially stable, inversely exponentially stable and of same order as $G_0$.

S3) The sequences $\{e_i(t)\}$ $i = 1,2$ consist of independent random variables with zero mean, variance $\sigma^2$ and finite fourth order moments.

C1) The controller $C$ is given by
$$C(q,\rho) = \frac{\rho_1 + \rho_3 q^{-1} + \ldots + \rho_{2n-1}q^{-n}}{1+\rho_2 q^{-1} + \rho_4 q^{-2} + \ldots + \rho_{2n}q^{-n}}$$

Assumptions S1), S2) and C1) imply that the optimal controller is given by (6).

We shall perform the analysis in a neighbourhood of the optimal controller. We therefore assume that the initial controller is the optimal controller

$$\rho_0 = \rho_{opt}$$

We then take one step in the Gauss-Newton direction, i.e.

$$\rho_1 = \rho_0 - \gamma_0\hat{R}_N^{-1}\hat{J}_N'(\rho_0) \qquad (11)$$

$$\hat{R}_N = \frac{1}{N}\sum_{t=1}^{N}\hat{y}'(t,\rho_0)[\hat{y}'(t,\rho_0)]^T \qquad (12)$$

with the step-size $\gamma_0 = 1$. The size of this step is a measure of the accuracy of the method around the optimum and in this section we shall analyze the properties of this step in the sequel.

### 4.1 *Accuracy in the parameter space*

The change[3]

$$\tilde{\rho}_N = \rho_1 - \rho_0 = \rho_1 - \rho_{opt}$$

in the parameter vector is non-zero even though the gradient $J'(\rho_0) = J'(\rho_{opt})$ is zero. The reason is that $\hat{J}'_N(\rho_0) = \hat{J}'_N(\rho_{opt})$ is a finite data approximation of the gradient that in addition is corrupted by the noise in the second experiment. A measure of this imprecision is the asymptotic covariance matrix

$$P = \lim_{N\to\infty} N\,\mathrm{E}[\tilde{\rho}_N\tilde{\rho}_N^T] \qquad (13)$$

Under the given assumption

$$y_1(t,\rho_{opt}) = e_1(t) \qquad (14)$$

$$\begin{aligned}\hat{y}'(t,\rho_{opt}) &= -C'(\rho_{opt})\left(\frac{G_0}{H_0}e_1(t) + \frac{G_0F^{-1}}{H_0-1}e_2(t)\right)\\ &\triangleq -C'(\rho_{opt})z_1(t) - C'(\rho_{opt})z_2(t)\\ &\triangleq \varphi_1(t) + \varphi_2(t) \qquad (15)\end{aligned}$$

where $\varphi_2$ is due to the noise in the second experiment.

Introduce $R_{ii} = \mathrm{E}[\varphi_i(t)\varphi_i^T(t)]$. Under S1)-S3) and C1)

$$R \triangleq \lim_{N\to\infty} \hat{R}_N = R_{11} + R_{22} \qquad \text{w.p. 1}$$

and

$$P = R^{-1}QR^{-1} \qquad (16)$$

where the variability $Q$ of the gradient estimate $\hat{J}'_N(\rho_{opt})$ is given by

$$\begin{aligned}Q &= \lim_{N\to\infty} N\mathrm{E}\left[\hat{J}'_N(\rho_{opt})\hat{J}'_N(\rho_{opt})^T\right]\\ &= \sigma^2\left(R_{11} + R_{22}\right)\end{aligned}$$

From this we see that the noise term $\varphi_2$ has both a benign and a malign effect: Its contribution to $R$ decreases the step and hence reduces the asymptotic parameter variance. This is a regularizing effect. However, its contribution to the variability of the gradient estimate increases the variance. Since (16) is quadratic in $R^{-1}$ the net effect will be that the variance is reduced. In fact

$$P = \sigma^2\left[R_{11} + R_{22}\right]^{-1}$$

However, the regularization term $R_{22}$ has an undesired side-effect, namely that the convergence *rate* will be slower since shorter steps are taken.

In order to get a meaningful analyzis we will from now on assume that the regularizing effect can be neglected, i.e. $R_{22}$ can be neglected in $R$, but that the contribution of $\varphi_2$ to the variability of the gradient estimate can not be neglected, i.e. $Q$ depends on $R_{22}$. Under this assumption we have

$$P = \sigma^2R_{11}^{-1} + \sigma^2R_{11}^{-1}R_{22}R_{11}^{-1} \qquad (17)$$

Equation (17) does not provide much insight into how the algorithm depends on the system parameters. In the next sub-section we provide frequency domain expressions which are much easier to interpret.

### 4.2 *Accuracy in the frequency domain*

Introduce

$$\Delta_N(e^{i\omega}) = C(e^{i\omega},\rho_1) - C(e^{i\omega},\rho_{opt})$$

When the error $\tilde{\rho}_N$ is small, which happens when $N$ is large, it is possible to make the following approximation

$$\Delta_N(e^{i\omega}) \approx C'(e^{i\omega},\rho_{opt})^T\tilde{\rho}_N \qquad (18)$$

where $C'(e^{i\omega},\rho_{opt}) = \frac{\partial C}{\partial\rho}C(e^{i\omega},\rho)|_{\rho=\rho_i}$. This gives

$$\begin{aligned}&\mathrm{E}[|\Delta_N(e^{i\omega})|^2] \approx\\ &C'(e^{i\omega},\rho_{opt})^T\mathrm{E}[\tilde{\rho}_N\tilde{\rho}_N^T]C'(e^{-i\omega},\rho_{opt})\end{aligned}$$

Under certain assumptions, the asymptotic variance expression in Theorem 3.1 in Ljung (1985) can be extended to cover this expression. With $\Phi_{ii}$ as the spectrum of signal $z_i$ in (15) we have the following result.

*Proposition 1.*

$$\begin{aligned}&\lim_{n\to\infty}\lim_{N\to\infty}\frac{N}{n}\mathrm{E}[|\Delta_N(e^{i\omega})|^2] =\\ &\sigma^2\frac{1}{\Phi_{11}(e^{i\omega})} + \sigma^2\frac{\Phi_{22}(e^{i\omega})}{\Phi_{11}^2(e^{i\omega})}\end{aligned}$$

For finite but large values of the system order $n$ and the number of data $N$, the implication of Proposition 1 is that

$$\begin{aligned}&\mathrm{E}[|\Delta_N(e^{i\omega})|^2] \approx\\ &\sigma^2\frac{n}{N}\left|\frac{H_0}{G_0}\right|^2\left(1 + \frac{1}{\left|(1-H_0^{-1})F\right|^2}\right)\end{aligned}$$

### 4.3 *Performance degradation*

Since the update step (11) takes us away from the optimal point, the performance of the closed loop will degrade, i.e.

$$J_V(\rho_1) \triangleq J(\rho_1) - J(\rho_{opt}) > 0$$

For full order minimum variance control, this difference can be written

$$J_V(\rho_1) = \mathrm{E}[(y(t,\rho_1) - y(t,\rho_{opt}))^2] \qquad (19)$$

[3] Subscript $N$ indicates that experiments of length $N$ have been performed.

In Hjalmarsson *et al.* (1996) (see also Gevers and Ljung (1986)) it is shown that for small $\bar{\rho}_N$ one may write

$$J_V(\rho_1) \approx \sigma_1^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{G_0(e^{i\omega})}{H_0(e^{i\omega})} \right|^2 \mathrm{E}[|\Delta_N(e^{i\omega})|^2] d\omega$$

Using this together with Proposition 1 gives that

$$J_V \approx J_v^{IFT} \triangleq \sigma^2 \frac{n}{N} \left(1 + W_1((1 - H_0^{-1})F)\right) \tag{20}$$

where

$$W_1(Z) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|Z(e^{i\omega})|^2} d\omega$$

*Remarks:*

- Notice that the expression (20) depends only on the noise dynamics $H_0$ and not at all on the system dynamics $G_0$.
- The second term is due to the noise in the second experiment.

## 5. A COMPARISION WITH OPTIMAL INPUT DESIGN

### 5.1 *System identification using optimal input design*

The minimum variance control problem on an unknown plant can also be solved by first performing an identification experiment on the plant, identifying the system and noise dynamics and then implementing the controller

$$\hat{C}_N = \frac{\hat{H}_N - 1}{\hat{G}_N}$$

where $\hat{G}_N$ and $\hat{H}_N$ are the estimates of the system $G_0$ and noise dynamics $H_0$, respectively. In Gevers and Ljung (1986) (see also Hjalmarsson *et al.* (1996)), the problem of designing the system identification experiment such that the performance degradation (19) is minimized is considered. The result is that the optimal choice of experimental conditions is to perform the identification in closed loop using the minimum variance controller in the loop. This gives the following result:

$$\mathrm{E}[|\Delta_N(e^{i\omega})|^2] \approx \frac{n}{N} \left| \frac{H_0}{G_0} \right|^2 \sigma^2 \tag{21}$$

$$J_V \approx J_v^{id} \triangleq \sigma^2 \frac{n}{N} \tag{22}$$

*Remark:* Notice that this scheme is infeasible since the system must be known in order to compute the minimum variance controller.

### 5.2 *Comparison*

Comparing the expression (20) for the performance degradation for the Gauss-Newton procedure with the corresponding expression (22) for the optimal system identification procedure we see that $J_v^{IFT} \geq J_v^{id}$. In fact equality will only hold if $e_2 \equiv 0$.

Comparing the performance degradation is not completely fair since IFT uses two experiments of length $N$ whereas the identification procedure only uses one experiment of length $N$. In order to make the comparison more fair we shall compare the two procedures when they use the same energy of the system (measured at the output).

In the system identification case, the output power $\mathrm{E}[y^2(\rho_{opt})] = \sigma^2$. Hence the energy used in the system identification case is $N_{id}\sigma^2$ where $N_{id}$ is the number of measurements that are collected from the system.

With

$$W_2(Z) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} |Z(e^{i\omega})|^2 d\omega$$

the power in the first IFT experiment is $\sigma^2$ and

$$\mathrm{E}[y_2^2(\rho_{opt})] = \sigma^2 \left(1 + W_2((1 - H_0^{-1})F)\right)$$

in the second experiment, Thus the total power is

$$\mathrm{E}[y_1^2(\rho_{opt}) + y_2^2(\rho_{opt})] = \sigma^2 \left(2 + W_2((1 - H_0^{-1})F)\right)$$

Equating the energies gives

$$N_{id} = N_{IFT} \left(2 + W_2((1 - H_0^{-1})F)\right)$$

where $N_{IFT}$ is the experiment length in the IFT approach. Now, the ratio $J_v^{IFT}/J_v^{id}$ when the energies are equal is given by

$$V = J_v^{IFT}/J_v^{id} = \left(1 + W_1((1 - H_0^{-1})F)\right) \times \left(2 + W_2((1 - H_0^{-1})F)\right) \tag{23}$$

From (23) it is obvious that the accuracy using system identification is better. One should however recall that an external reference signal is required in the system identification in order to make the system identifiable. Hence, in practice the difference will be less than the expression (23) suggests.

When no prefilter is used, i.e. $F \equiv 1$, we see from (23) that a particularly difficult case for IFT is when $H_0^{-1} \approx 1$ on the unit circle since then $W_1$ becomes large. But this corresponds to the case where the open loop disturbance is almost white and feedback can do very little to improve performance.

### 5.3 *Choosing prefilter in IFT*

The prefilter $F$ is used to prefilter the output of the first experiment before it is applied as reference signal to the second experiment: $r(t) = F(q)y_1(t, \rho_i)$. The accuracy will be influenced by this prefilter. Minimizing the performance degradation w.r.t. $F$ under the constraint that the total

power should be equal corresponds to minimizing $V$ in (23) w.r.t. $F$. The following result holds.

*Proposition 2.* The ratio (23) is minimized by the choice

$$F = \sqrt{2}\frac{1}{1 - H_0^{-1}} \tag{24}$$

and $\min_F V = (1 + \frac{1}{\sqrt{2}})(1 + \sqrt{(2)}) \approx 5.8$.

PROOF. See Appendix A. □

*Remarks:*

- With the choice (24), the ratio (23) is independent of both the system dynamics and the noise spectrum.
- The gain in accuracy using prefilter is dramatic when $H_0$ is close to unity.
- The optimal prefilter depends on the noise spectrum which is unknown. Notice, however, that the output in the second experiment is white when the optimal prefilter is used. This can be used as guideline when selecting a prefilter.

## 6. CONCLUSIONS

Frequency domain expressions for the accuracy of Iterative Feedback Tuning have been derived in the case of high order minimum variance control. The expressions show that the accuracy does not depend on the system dynamics but only on the noise spectrum. Furthermore, an optimal prefilter (24) was derived. A comparison with system identification, where optimal input design was been used, has also been presented. It was shown that optimal input design yields more than five times higher accuracy than IFT. Notice, however, that the comparison was made at the optimum conditions – the optimal controller was already in the loop. A more realistic comparison would be to assume suboptimal operating conditions. This is a topic for future research.

## 7. REFERENCES


Åström, K.J. and B. Wittenmark (1984). *Computer Controlled Systems.* Prentice-Hall. Englewood Cliffs, New Jersey.

Benveniste, A., M. Métivier and P. Priouret (1987). *Adaptive Algorithms and Stochastic Approximations.* Springer-Verlag. Berlin.

Gevers, M. and L. Ljung (1986). Optimal experiment designs with respect to the intended model application. *Automatica* **22**, 543–554.

Hjalmarsson, H., M. Gevers and F. De Bruyne (1996). For some model based control design criteria, closed loop identification gives better performance. *Automatica* **32**, 1659–1673.

Hjalmarsson, H., S. Gunnarsson and M. Gevers (1994). A convergent iterative restricted complexity control design scheme. In: *Proc. 33rd IEEE CDC.* Orlando, Florida. pp. 1735–1740.

Lee, W.S., B.D.O. Anderson, R.L. Kosut and I.M.Y Mareels (1993). A new approach to adaptive robust control. *Int. Journal of Adaptive Control and Signal Processing* **7**, 183–211.

Ljung, L. (1985). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Automatic Control* **AC-30**, 834–844.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall. Englewood Cliffs, NJ.

Robbins, H. and S. Monro (1951). A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407.

Schrama, R.J.P. (1992). Approximate Identification and Control Design. PhD thesis. Delft University of Technology.

Zang, Z., R.R. Bitmead and M. Gevers (1995). Iterative weighted least-squares identification and weighted LQG control design. *Automatica* **31**, 1577–1594.


## Appendix A. PROOF OF PROPOSITION 2

Consider first the problem

$$\begin{aligned} &\min_{W_1,\, W_2} \quad (1 + W_1)(2 + W_2) \\ &\text{Subject to} \quad W_1 W_2 \geq 1,\ W_1 \geq 0 \end{aligned} \tag{A.1}$$

For fixed $W_2$ this function is minimized by $W_1 = 1/W_2$. Hence the solution to (A.1) is obtained by finding the minimum to

$$(1 + 1/W_2)(2 + W_2)$$

which is $W_2 = \sqrt{2}$. The optimum choice of $W_1$ is $W_1 = \sqrt{2}$ and the corresponding minimum of the function is $(1 + \frac{1}{\sqrt{2}})(2 + \sqrt{2})$.

Consider now the problem $\min_F V$. Taking

$$F = \sqrt{2}\frac{1}{1 - H_0^{-1}} \tag{A.2}$$

gives $V = (1+\frac{1}{\sqrt{2}})(2+\sqrt{2})$. Since $\min_F V$ involves the same criterion as (A.1) but the constraint on $W_1$ and $W_2$ in (A.1) is a subset of the constraints in $\min_F V$ this must be the optimum solution. That the constraint $W_1 W_2 \geq 1$ is included in the problem $\min_F V$ follows from Cauchy-Schwarz inequality

$$1 \leq W_1((1 - H_0^{-1})F)W_2((1 - H_0^{-1})F)$$