Pergamon

0005-1098(94)00112-X

*Automatica*, Vol. 31, No. 3, pp. 367-379, 1995
Elsevier Science Ltd
Printed in Great Britain
0005-1098/95 $9.50 + 0.00

# Optimum Realizations of Sampled-data Controllers for FWL Sensitivity Minimization*

ANTON G. MADIEVSKI,† BRIAN D. O. ANDERSON† and MICHEL GEVERS‡

*A procedure for choosing a realization of a digital compensator of known transfer function is described, which ensures that the errors introduced into a sampled-data closed loop by using finite-word-length arithmetic in the compensator operation are minimized.*

**Key Words**—Closed-loop systems; control systems; convergence of numerical methods; digital control; linear systems; numerical methods; sampled data systems; time-varying systems.

**Abstract**—The problem of an optimal finite-word-length state-space realization of a digital controller is investigated. The closed loop to be considered consists of a continuous-time plant, a discrete-time controller, a sampler, a zero-order hold and an antialiasing filter. An effective algorithm is proposed to find the optimal sampled-data controller realization minimizing the sensitivity of the closed-loop performance with respect to coefficient errors in the state variable matrices of the controller realization. In order to get a tractable problem, a two-step procedure is to be used: very fast sampling at a multiple of the sampling frequency followed by 'blocking' or 'lifting' to obtain a single-rate system. The procedure allows consideration of the system's intersample behaviour.

## 1. INTRODUCTION

In order to achieve the desired characteristics of a closed-loop system, a controller is to be used. It is well-known that a desired controller's transfer function can be implemented by any one of an infinite set of realizations of the controller. Though all these realizations are in principle equivalent, since they yield the same transfer function, they have different numerical properties due to finite-word-length effects when they are implemented by a digital device. Such factors as sensitivity and error propagation strongly affect closed-loop performance, and are responsible for differences between desired ideal closed-loop characteristics and those actually obtained. A problem of great importance is to find the realization of the controller that achieves the best performance of the closed-loop system, i.e. that gives the best approximation of the ideal closed-loop behaviour.

Results on optimal realizations of filters (or 'open-loop systems') minimizing some measure of performance degradation due to FWL errors date back to the late 1970s. The first results were on realizations that minimize roundoff error propagation (Hwang, 1977; Mullis and Roberts, 1976). Realizations minimizing some measure of the transfer function sensitivity to coefficient errors took much longer to emerge (Thiele, 1986).

It was not until the late 1980s that the problem of optimal controller realization minimizing closed-loop performance degradation due to numerical errors was addressed. Solutions were proposed first for specific control schemes (LQG, pole placement), and more recently for general two-degree-of-freedom controllers (Li and Gevers, 1990a, b, 1991; Liu and Skelton, 1990; Liu *et al.*, 1992; Williamson and Kadiman, 1989). The last three references provide an optimal FWL-LQG design (which includes an optimal realization in the design process). Liu and Skelton (1990) and Liu *et al.* (1992) provide an optimal approach, while Williamson and Kadiman (1989) provide a suboptimal approach.

A survey of these results can be found in

* Received 7 March 1994; revised 9 June 1994; received in final form 19 July 1994. The original version of this paper was presented at the 12th IFAC World Congress, which was held in Sydney, Australia during 19–23 July 1993. The Published Proceedings of this IFAC Meeting may be ordered from: Elsevier Science Limited, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, U.K. This paper was recommended for publication in revised form by Associate Editor K. Uchida under the direction of Editor T. Başar. Corresponding author Dr Anton Madievski. Tel (61) 249 4581; Fax (61) 6 249 2698;
E-mail ant101@syseng.anu.edu.au.

† Department of Systems Engineering, Australian National University, Canberra, ACT 0200, Australia.

‡ Centre d'Ingéneirie des Systèmes, d'Automatique et de Mécanique Appliquée, Université Catholique de Louvain, Bâtiment Euler, 4–6, avenue Georges Lemaître, B-1348 Louvain-la Neuve, Belgique.

Gevers and Li (1993). The methods essentially differ in the choice of performance measure (either roundoff error propagation or transfer function sensitivity) and in the norms used to evaluate this performance degradation. In Gevers and Li (1993) a synthetic measure of performance degradation of a closed-loop system, incorporating both roundoff errors and coefficient errors, was minimized with respect to all compensator realizations. The results on closed-loop sensitivity minimization in the above references all pertain to sensitivity measures of the closed-loop transfer function with respect to controller parameter errors. In Li and Gevers (1993) a weighted sensitivity measure of the *closed-loop poles* with respect to controller parameter errors is minimized.

The common feature of all these optimal controller realization results is that the system to be controller is assumed to be described by a discrete-time transfer function $H(z)$. In most practical applications a digital controller is used to control a continuous-time plant, using both a sampler and a hold device.

Any optimization using solely a discrete-time transfer function of the closed loop neglects the intersample system behaviour and particularly intersample ripple. The novel contribution of this paper is to pose and solve a discrete-time compensator realization problem for a continuous–discrete closed-loop system, in which the digital controller acts on the continuous-time plant via a zero-order hold device, and in which the tracking error of the continuous system is passed through an antialiasing filter and then sampled. With this continuous–discrete set-up, the performance measure involves, of necessity, a hybrid operator: it is a measure of the sensitivity of the closed-loop input–output operator to the parameters of the compensator realization.

The outline of this paper is as follows: in Section 2 we establish the definitions of sensitivity 'functions' (operators) and the $\mathscr{L}_2$ sensitivity measure of a closed-loop system. In Section 3 we study the finite-word-length optimal realization minimizing a measure of the sensitivity of the closed-loop operation with respect to controller coefficient errors. (No claim is made about FWL roundoff noise effects.) The existence and uniqueness of an optimal solution are established. A recursive algorithm for obtaining the optimal solution is given. A two-step procedure (fast sampling followed by blocking) that allows consideration of intersample behaviour of a closed-loop system is described and studied in Section 4. Two numerical examples to confirm theoretical results are given in Section 5, followed by some concluding remarks in Section 6.

## 2. SENSITIVITY MEASURE OF A REALIZATION

First consider a discrete linear time-invariant multi-input, multi-ouput controller having a transfer function $K(z)$ that can be expressed in terms of matrices $A$, $B$, $C$ and $D$ of a minimal state-space realization as follows:

$$K(z) = C(zI_R - A)^{-1}B + D, \tag{1}$$

where $A \in \mathbb{R}^{R\times R}$, $B \in \mathbb{R}^{R\times L}$, $C \in \mathbb{R}^{M\times R}$, $D \in \mathbb{R}^{M\times L}$ and $K \in \mathbb{C}^{M\times L}$. Clearly, if the matrices $A$, $B$, $C$ and $D$ satisfy (1) then, for any similarity transformation $T$, the matrices $T^{-1}AT$, $T^{-1}B$, $CT$ and $D$ also satisfy (1). This means that there exist an infinite number of representations of the system. All these representations are equivalent insofar as they yield the same transfer function. However, different realizations have different numerical properties such as sensitivity to coefficient errors and propagation of signal roundoff errors. This means that in the finite-precision case all these realizations are no longer equivalent. In practice it is impossible to realize the matrices $A$, $B$, $C$ and $D$ exactly owing to finite-word-length (FWL) constraints. As a result, the transfer function given by (1) and the transfer function with the matrices $A$, $B$, $C$ and $D$ replaced by their FWL versions are different. Since different FWL realizations have different sensitivities, our task is to search for those realizations that minimize the sensitivity in some appropriate measure reflecting the overall control objective.

In order to define such a measure, we shall use the derivatives of elements of the controller transfer function matrix at an arbitrary but fixed value of $z$ with respect to the elements of the matrices $A$, $B$ and $C$ of the realization:

$$\frac{\partial k_{m,l}}{\partial a_{r,q}} = g_{m,r}f_{q,l}, \quad \frac{\partial k_{m,l}}{\partial b_{r,l}} = g_{m,r}\delta_{l,i}, \quad \frac{\partial k_{m,l}}{\partial c_{j,r}} = f_{r,l}\delta_{m,j}, \tag{2}$$

where $a$, $b$, $c$, $k$, $g$ and $f$ are elements of the matrices $A$, $B$, $C$, $K$, $G$ and $F$, respectively, with

$$G = C(zI_R - A)^{-1} \in \mathbb{C}^{M\times R}, \tag{3a}$$

$$F = (zI_R - A)^{-1}B \in \mathbb{C}^{R\times L}, \tag{3b}$$

$l, i = 1, 2, \ldots, L$, $m, j = 1, 2, \ldots, M$, $r, q = 1, 2, \ldots, R$ and $\delta$ is the Kronecker delta. Note that the matrix $D$ is coordinate-independent and

has nothing to do with the optimal realization problem.

Our major goal is to find the optimal implementation of the controller for achieving the best performance of the closed-loop system where the controller is implemented with FWL. 'Best performance' can mean many things. As made more precise below, we shall consider the accuracy of implementing the input–output operator for the closed loop.

Consider a hybrid closed loop where the plant is continuous-time and the controller is discrete-time (such a configuration represents the usual situation). This closed loop is drawn in Fig. 1, where $\Pi$ stands for the $L \times M$ continuous-time plant, $K$ for the $M \times L$ discrete controller, $\Phi$ for the strictly proper stable antialiasing filter, $\Sigma$ for the sampler with the sampling period $\tau$ and $H$ for the hold element, here assumed to be a zero-order hold. (In the multivariable situation $\Phi$, $\Sigma$ and $H$ are diagonal operators.)

First of all, we need to define a sensitivity measure of the closed-loop operator with respect to errors in the realization $A$, $B$ and $C$ of the controller. Then the problem of minimizing of this measure will arise.

Earlier works (Gevers and Li, 1993; Li and Gevers, 1990a, b, 1991, 1993; Liu and Skelton, 1990; Liu *et al.*, 1992; Williamson and Kadiman, 1989) looked at purely discrete-time problems, and it was possible (and easier) to deal just with frequency-domain quantities. However, in order to take into consideration intersample behaviour of the hybrid closed loop, we need to work in the time domain.

### 2.1. *Closed-loop operator and sensitivity with respect to a controller parameter*

We shall assume that the sampling interval is such that no unstabilizable or undetectable modes are introduced by the sampling operator and that the closed loop is stable. Unstabilizable or undetectable modes can only occur for nongeneric $\Pi(s)$, and, even then, only for isolated choices of sampling interval (Francis and Georgiou, 1988). Stability of the closed-loop system means that with zero input, any nonzero initial state decays to zero exponentially fast, and $u(\cdot) \in L_p^L[0, \infty)$ implies $y(\cdot) \in L_p^L[0, \infty)$ for all $p \in [1, \infty]$ (see Francis and Georgiou, 1988). The closed loop is defined by a linear periodically time-varying operator $\mathscr{X}$ with associated causal impulse response $\mathscr{X}(t, s)$, such that

$$\bar{y}(t) = \int_{-\infty}^{t} \mathscr{X}(t, s)\bar{u}(s)\, ds, \tag{4}$$

$$\mathscr{X}(t + \tau, s + \tau) = \mathscr{X}(t, s). \tag{5}$$

The stability condition is expressed by

$$\|\mathscr{X}(t, s)\|_F \leq a \exp[-\beta(t - s)], \tag{6}$$

for some $a > 0$ and $\beta > 0$, with the subscript F denoting the Frobenius norm, i.e.

$$\|A\|_F = [\mathrm{tr}\,(A^T A)]^{1/2}.$$

A composition of two stable operators is stable.

Formally, with minimal abuse of notation, we can write

$$\mathscr{X} = \Pi H K \Sigma \Phi (I + \Pi H K \Sigma \Phi)^{-1}, \tag{7}$$

where $\Pi$, $H$, $K$, $\Sigma$, and $\Phi$ are the operators corresponding to the blocks shown in Fig. 1, and then the derivative of $\mathscr{X}$ with respect to an element in a realization of $K$ can be formally written as

$$\frac{\partial \mathscr{X}}{\partial a} = \mathscr{V} \frac{\partial K}{\partial a} \mathscr{W}, \tag{8a}$$

where

$$\mathscr{V} = (I_L + \Pi H K \Sigma \Phi)^{-1} \Pi H, \tag{8b}$$

$$\mathscr{W} = \Sigma \Phi (I_L + \Pi H K \Sigma \Phi)^{-1}. \tag{8c}$$

Because of the stability of the closed loop, $\mathscr{V}$ and $\mathscr{W}$ map $l_p^M(\mathbb{Z}_+)$ into $\mathscr{L}_p^L[0, \infty)$ and $\mathscr{L}_p^L[0, \infty)$ into $l_p^L(\mathbb{Z}_+)$, respectively, for all $p \in [1, \infty]$, including of course $p = 2$. Moreover, the mappings are causal.

The derivative (8a) of the closed-loop operator $\mathscr{X}$ can be represented as in Fig. 2.

The representation of Fig. 2 has a deficiency that should be remedied. If $K(z)$ is open-loop unstable then it is $\partial K(z)/\partial a$ (see (2) and (3)); yet if the closed loop is stable, one would expect that the operator $\partial \mathscr{X}/\partial a$ should also have this property. This is in fact so. The representation of Fig. 2 can be replaced by one (with lower overall state dimension) that is stable. This is done as follows.

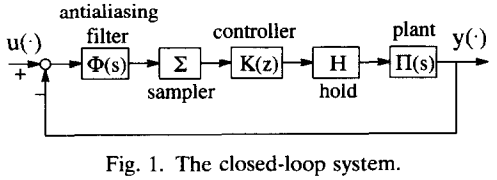


Fig. 1. The closed-loop system.

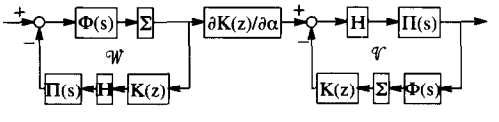


Fig. 2. Representation of the derivative of the closed-loop operator with respect to the parameter $\alpha$ of the controller.

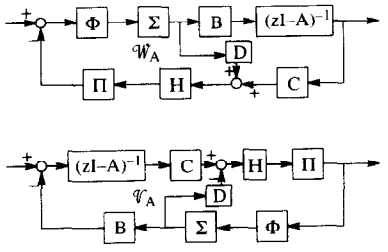


Fig. 3. Modification of the operators $\mathcal{V}$ and $\mathcal{W}$ used in the alternative construction of the derivatives of the closed-loop operator.

We note that if $a = a_{i,j}$ then (with some minor abuse of notation)

$$\frac{\partial \mathcal{X}}{\partial a_{i,j}} = (I_L + \Pi HK\Sigma\Phi)^{-1}\Pi HC(zI_R - A)^{-1}e_i e_j^{\mathrm{T}} \times (zI_R - A)^{-1}B\Sigma\Phi(I_L + \Pi HK\Sigma\Phi)^{-1} = \mathcal{V}_A e_i e_j^{\mathrm{T}} \mathcal{W}_A, \tag{9a}$$

where $\mathcal{V}_A$ and $\mathcal{W}_A$ (depicted in Fig. 3) are stable operators differing marginally from $\mathcal{V}$ and $\mathcal{W}$ in terms of the points where the loop input is introduced or the loop output is taken from. Also, the range of $\mathcal{W}_A$ and the domain of $\mathcal{V}_A$ are discrete-time signals: $\mathcal{W}_A$ and $\mathcal{V}_A$ are bounded operators from $L_p^{\mathrm{L}}[0, \infty)$ to $l_p^{\mathrm{R}}(\mathbb{Z}_+)$ and $l_p^{\mathrm{R}}(\mathbb{Z}_+)$ to $L_p^{\mathrm{L}}[0, \infty)$ $\forall p \in [1, \infty]$. Similarly

$$\frac{\partial \mathcal{X}}{\partial b_{i,j}} = \mathcal{V}_A e_i e_j^{\mathrm{T}} \mathcal{W}, \tag{9b}$$

$$\frac{\partial \mathcal{X}}{\partial c_{i,j}} = \mathcal{V} e_i e_j^{\mathrm{T}} \mathcal{W}_A. \tag{9c}$$

Since $\mathcal{V}$, $\mathcal{W}$, $\mathcal{V}_A$ and $\mathcal{W}_A$ are all stable operators, the operators on the right in (9) are all stable.

In formulating a sensitivity 'function' (here, more properly, an operator) associated with the realization of a system, it is conventional to organize the matrix calculations slightly differently; one picks a particular entry of $\mathcal{X}$, $\mathcal{X}_{k,l}$ say, and constructs the matrices $\partial \mathcal{X}_{k,l}/\partial A$, $\partial \mathcal{X}_{k,l}/\partial B$ and $\partial \mathcal{X}_{k,l}/\partial C$, where the $(i, j)$ entry of $\partial \mathcal{X}_{k,l}/\partial A$ is $\partial \mathcal{X}_{k,l}/\partial a_{i,j}$. In the light of (9), it is clear that

$$\frac{\partial \mathcal{X}_{k,l}}{\partial A} = \mathcal{V}_A^{\mathrm{T}} e_k e_l^{\mathrm{T}} \mathcal{W}_A^{\mathrm{T}}, \tag{10a}$$

$$\frac{\partial \mathcal{X}_{k,l}}{\partial B} = \mathcal{V}_A^{\mathrm{T}} e_k e_l^{\mathrm{T}} \mathcal{W}_A^{\mathrm{T}}, \tag{10b}$$

$$\frac{\partial \mathcal{X}_{k,l}}{\partial C} = \mathcal{V}_A^{\mathrm{T}} e_k e_l^{\mathrm{T}} \mathcal{W}_A^{\mathrm{T}}. \tag{10c}$$

Here $\mathcal{V}^{\mathrm{T}}$ is not the adjoint operator of $\mathcal{V}$; rather, $\mathcal{V}^{\mathrm{T}}$ is defined by the condition $e_i^{\mathrm{T}} \mathcal{V}^{\mathrm{T}} e_j = e_j^{\mathrm{T}} \mathcal{V} e_i =$ $(j, i)$ component of $\mathcal{V}$; thus $\mathcal{V}^{\mathrm{T}}$ is $\mathcal{V}$ with elements reorganized.

Let us sum up our results to this point.

*Theorem* 2.1. Consider the closed-loop system depicted in Fig. 1, comprising a strictly proper (nonzero) stable antialiasing filter $\Phi(s)$, a sample $\Sigma$ with sampling interval $\tau$, a (nonzero) discrete-time controller $K(z)$, a zero-order hold $H$ and a (nonzero) plant $\Pi(s)$. Suppose that $\tau$ is such that no unstabilizable or undetectable modes are introduced by the sampling operator, and the closed loop, defined by a periodically time-varying impulse response $\mathcal{X}(t, s)$, is stable. Let the controller have a minimal state-variable realization $C(zI - A)^{-1}B + D$, and let stable closed-loop operators $\mathcal{V}$, $\mathcal{W}$, $\mathcal{V}_A$ and $\mathcal{W}_A$ be defined as depicted in Figs 2 and 3. Then the sensitivity functions of $\mathcal{X}$ with respect to the elements of $A$, $B$, $C$ in the controller realization are given by (10) for $k, l = 1, 2, \ldots, L$.

## 2.2. *A numerical sensitivity measure*

In order to determine a single numerical measure of sensitivity, we shall use a norm associated with the sensitivity 'functions'. This is not an induced norm, but rather a norm associated with the impulse response representation of a stable operator, viewing it simply as a function of time in two variables. We confine attention to a matrix impulse response $\mathcal{U}(t, s)$ defined in the half-plane $s \le t$, with the periodicity property $\mathcal{U}(t + \tau, s + \tau) = \mathcal{U}(t, s)$ and with the (exponential) stability property

$$\|\mathcal{U}(t, s)\|_{\mathrm{F}} \le \alpha \exp[-\beta(t - s)]$$

for some positive, $\alpha, \beta$. (We remark that in Francis and Georgiou (1988) there is also a departure from the case of induced norms in defining norms for a stable, periodically time-varying linear system.) The norm is

$$\|\mathcal{U}\|_2 = \left[\int_0^\tau \mathrm{d}t \int_{-\infty}^{t} \|\mathcal{U}(t, s)\|_{\mathrm{F}}^2 \,\mathrm{d}s\right]^{1/2}. \tag{11}$$

Notice that, in view of the periodicity property, the norm reflects all values assumed by $\mathcal{U}(t, s)$ in $-\infty < s \le t < \infty$, even though the integration with respect to $t$ only extends over $[0, \tau]$. Note also the alternative expression obtained by changing the order of integration:

$$\|\mathcal{U}\|_2 = \left[\int_0^\tau \mathrm{d}s \int_{s}^{\infty} \|\mathcal{U}(t, s)\|_{\mathrm{F}}^2 \,\mathrm{d}t\right]^{1/2}. \tag{12}$$

Now we can define the sensitivity measures of the closed-loop operator with respect to the realization $A$, $B$, and $C$ of the controller.

*Definition* 2.1. The sensitivity measure $M_2$ of the closed-loop operator with respect to the realization of the controller is the sum of the squares of the $\mathscr{L}_2$ norms of sensitivity operators of the closed loop with respect to the matrices $A$, $B$ and $C$ of the realization of the controller:

$$M_2 = \sum_{k,l} \left( \left\| \frac{\partial \mathscr{H}_{k,l}}{\partial A} \right\|_2^2 + \left\| \frac{\partial \mathscr{H}_{k,l}}{\partial B} \right\|_2^2 + \left\| \frac{\partial \mathscr{H}_{k,l}}{\partial C} \right\|_2^2 \right). \quad (13)$$

It is worth noting how the measure differs from those employed in earlier optimum realization problems. First, the measure is intrinsically a time-domain rather than a frequency-domain one. Secondly, for reasons of mainly analytic convenience, in most earlier optimum realization work frequency-domain $\mathscr{L}_2$ norms of sensitivity functions related to $B$ and $C$ were used, while a frequency-domain $\mathscr{L}_1$ norm was used for the sensitivity function related to $A$. The frequency-domain $\mathscr{L}_2$ norms have some parallel (through Parseval's Theorem) with our time-domain $\mathscr{L}_2$ norms. Frequency-domain $\mathscr{L}_1$ norms of course are virtually unrelatable to a time-domain norm. Gevers and Li (1993) and Li and Gevers (1991) use a frequency-domain $\mathscr{L}_2$ norm for the sensitivity function related to $A$, and are closest in spirit to the present work, even though those authors use a discrete-time model for the system. Perkins *et al.* (1990) also uses an $\mathscr{L}_2$ norm in an optimal filter realization problem.

## 3. OPTIMAL FWL REALIZATIONS

The numerical measures of sensitivity defined above depend on the particular realization of the controller. We shall now clarify the nature of the dependence.

A coordinate-basis transformation $T$ transforms $(A, B, C)$ into $(T^{-1}AT, T^{-1}B, CT)$ and $(F, G)$ into $(T^{-1}F, GT)$. The operators $\mathscr{V}$ and $\mathscr{W}$ are unaltered, while $(\mathscr{V}_A, \mathscr{W}_A) \rightarrow (\mathscr{V}_A T, T^{-1}\mathscr{W}_A)$. Noting (10), we conclude that the sensitivity operators transform according to

$$\frac{\partial \mathscr{H}_{k,l}}{\partial A} \rightarrow T^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}}{\partial A} T^{-\mathrm{T}}, \quad (14a)$$

$$\frac{\partial \mathscr{H}_{k,l}}{\partial B} \rightarrow T^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}}{\partial B}, \quad (14b)$$

$$\frac{\partial \mathscr{H}_{k,l}}{\partial C} \rightarrow \frac{\partial \mathscr{H}_{k,l}}{\partial C} T^{-1}. \quad (14c)$$

Parenthetically, one can note that the corresponding formulae associated with (9) are not so attractive.

Let us make the definitions

$$J_{k,l}^B = \int_0^\tau \mathrm{d}t \int_{-\infty}^t \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial B} \left[ \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial B} \right]^{\mathrm{T}} \mathrm{d}s, \quad (15a)$$

$$J_{k,l}^C = \int_0^\tau \mathrm{d}t \int_{-\infty}^t \left[ \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial C} \right]^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial C} \mathrm{d}s. \quad (15b)$$

Evidently, under the coordinate-basis change,

$$J_{k,l}^B \rightarrow T^{\mathrm{T}} J_{k,l}^B T, \quad (16a)$$

$$J_{k,l}^C \rightarrow T^{-1} J_{k,l}^C T^{-\mathrm{T}}. \quad (16b)$$

We make the further definitions

$$J_B = \sum_{k,l} J_{k,l}^B, \quad (17a)$$

$$J_C = \sum_{k,l} J_{k,l}^C. \quad (17b)$$

Then, the second and third terms of the measure $M_2$ are precisely $\mathrm{tr}\, J_B$ and $\mathrm{tr}\, J_C$. Further, denoting the value of the measure $M_2$ after the coordinate-basis transformation by $M_2(T)$, we see that

$$M_2(T) = \sum_{k,l} \left\| T^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}}{\partial A} T^{-\mathrm{T}} \right\|_2^2 + \mathrm{tr}\,(J_B P) + \mathrm{tr}\,(J_C P^{-1}), \quad (18)$$

where

$$P = TT^{\mathrm{T}}. \quad (19)$$

It remains to consider in more detail how the first term in $M_2$ transforms when there is a coordinate-basis change. Notice that

$$\begin{aligned}
&\sum_{k,l} \left\| T^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}}{\partial A} T^{-\mathrm{T}} \right\|_2^2 \\
&= \sum_{k,l} \int_0^\tau \mathrm{d}t \int_{-\infty}^t \left\| T^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} T^{-\mathrm{T}} \right\|_{\mathrm{F}}^2 \mathrm{d}s \\
&= \sum_{k,l} \int_0^\tau \mathrm{d}t \int_{-\infty}^t \mathrm{tr} \left\{ T^{-1} \left[ \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} \right]^{\mathrm{T}} \right. \\
&\quad \left. \times TT^{\mathrm{T}} \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} T^{-\mathrm{T}} \right\} \mathrm{d}s \\
&= \sum_{k,l} \mathrm{tr} \int_0^\tau \mathrm{d}t \int_{-\infty}^t \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} T^{-\mathrm{T}} T^{-1} \\
&\quad \times \left[ \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} \right]^{\mathrm{T}} TT^{\mathrm{T}} \mathrm{d}s \\
&= \sum_{k,l} \mathrm{tr} \int_0^\tau \mathrm{d}t \int_{-\infty}^t \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} P^{-1} \left[ \frac{\partial \mathscr{H}_{k,l}(t,s)}{\partial A} \right]^{\mathrm{T}} P \mathrm{d}s.
\end{aligned} \quad (20)$$

Hence we can regard $M_2$ as a function of $P$:

$$M_2(P) = \mathrm{tr}\,(J_B P) + \mathrm{tr}\,(J_C P^{-1}) + \sum_k \sum_l \mathrm{tr} \int_0^\tau dt \int_{-\infty}^t \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} P^{-1} \times \left[\frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A}\right]^T P\, ds. \quad (21)$$

The optimal FWL compensator design can thus be formulated as follows:

$$P_{\mathrm{opt}} = \arg\min_{P>0} M_2(P). \quad (22)$$

If a solution exists then any square root $T_{\mathrm{opt}}$ such that $P_{\mathrm{opt}} = T_{\mathrm{opt}} T_{\mathrm{opt}}^T$ defines an optimal coordinate basis for the controller. We have an expression for the gradient of $P$:

$$\frac{\partial M_2(P)}{\partial P} = J_B - P^{-1} J_C P^{-1} + \sum_k \sum_l \int_0^\tau dt \int_{-\infty}^t \left\{ \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} P^{-1} \times \left[\frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A}\right]^T - P^{-1} \left[\frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A}\right]^T \times P \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} P^{-1} \right\} ds. \quad (23)$$

For evaluation purposes, the following formula is valuable; it is derived using standard properties of the Kronecker product and the function vec defined by Neudecker (1960) and proved in a number of textbooks and papers (e.g. Brewer, 1978):

$$\frac{\partial M_2(P)}{\partial P} = J_B - P^{-1} J_C P^{-1} + \sum_k \sum_l \mathrm{vec}^{-1} \left\{ \left[ \int_0^\tau dt \int_{-\infty}^t \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} \otimes \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} ds \right] \mathrm{vec}\, P^{-1} \right\} - \sum_k \sum_l \mathrm{vec}^{-1} \left\{ (P^{-1} \otimes P^{-1}) \times \left[ \int_0^\tau dt \int_{-\infty}^t \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} \otimes \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} ds \right]^T \mathrm{vec}\, P \right\}. \quad (24)$$

In the absence of an analytically computable value of $P$ producing a zero value of the gradient, a value $P_{\mathrm{opt}}$ of $P$ minimizing $M_2$ could be sought by an iterative algorithm

$$P_{i+1} = P_i - \mu \left. \frac{\partial M_2(P)}{\partial P} \right|_{P=P_i}, \quad (25)$$

where $\mu$ is a small positive number. The utility of the gradient algorithm is partly justified by the following theorem.

*Theorem* 3.1. Adopt the same hypothesis as in Theorem 2.1, and let $(A, B, C, D)$ be an initial realization of the controller $K(z)$. Let $J_B$ and $J_C$ be defined by (15)–(17), and let $M_2(P)$ define the sensitivity measure (21) of a controller realization obtained by transforming the initial realization through a nonsingular $T$, with $P = TT^T$. Then there exists a unique $P_{\mathrm{opt}} > 0$ that minimizes $M_2(P)$, and which accordingly can be found via an iterative gradient descent algorithm. (There always exists such small positive values of $\mu$ that guarantee the convergence.)

*Proof.* See the Appendix.

Once $P_{\mathrm{opt}}$ has been found, any square $T$ satisfying $TT^T = P_{\mathrm{opt}}$ can be selected. This defines $T$ to within right multiplication by an orthogonal matrix, and this additional freedom, present also in an all earlier optimal realization problems, can be exploited to force zero or unity entries into parts of $A$, $B$ or $C$ (see e.g. Li *et al.*, 1992); this has a beneficial practical effect, since obviously a zero or unity multiplication is realizable with no error.

## 4. EVALUATION OF THE SENSITIVITY MEASURE AND ITS GRADIENT

Now, in order to implement the iterative algorithm (25), we need to calculate the value of the gradient $\partial M_2(P)/\partial P$ at every iteration step. The problem is to calculate the values of

$$\int_0^\tau dt \int_{-\infty}^t \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} \otimes \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial A} ds, \quad (26a)$$

$$\int_0^\tau dt \int_{-\infty}^t \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial B} \left[\frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial B}\right]^T ds, \quad (26b)$$

$$\int_0^\tau dt \int_{-\infty}^t \left[\frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial C}\right]^T \frac{\partial \mathscr{X}_{k,l}(t,s)}{\partial C} ds. \quad (26c)$$

The prime concern of this section is to obtain a numerical procedure for calculating the three values above using standard techniques, i.e. standard software.

Figure 4 depicts the operators $\partial \mathscr{X}_{k,l}/\partial a_{i,j}$, $\partial \mathscr{X}_{k,l}/\partial b_{i,j}$ and $\partial \mathscr{X}_{k,l}/\partial c_{i,j}$. To understand these figures, recognize from (9) that

$$\frac{\partial \mathscr{X}_{k,l}}{\partial a_{i,j}} = (e_k^T \mathscr{V}_A e_i)(e_j^T \mathscr{W}_A e_l), \quad (27a)$$

$$\frac{\partial \mathscr{X}_{k,l}}{\partial b_{i,j}} = (e_k^T \mathscr{V}_A e_i)(e_j^T \mathscr{W} e_l), \quad (27b)$$

$$\frac{\partial \mathscr{X}_{k,l}}{\partial c_{i,j}} = (e_k^T \mathscr{V} e_i)(e_j^T \mathscr{W} e_l). \quad (27c)$$

The structure of this hybrid feedback loop is illustrated in Fig. 5, where the form of $\Delta_{da}(z)$

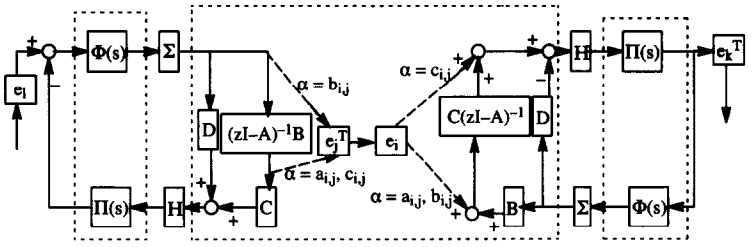


Fig. 4. The operator $\partial \mathscr{X}_{k,l}/\partial\alpha$ for various $\alpha$.

depends on the particular $a$. Note the implicit definitions of $\Delta_{da}(z)$ and $\Omega(s)$.

Because of the mixture of continuous- and discrete-time entities, some of the mappings in the hybrid system are operators that have no transfer function representation, and thus the operators $\mathscr{V}_A$ and $\mathscr{W}_A$ do not have transfer function representations. In order to facilitate the calculations of the $\mathscr{L}_2$ norms, the continuous-time part of the hybrid system is approximated by a discrete-time system with arbitrarily fast sampling. This can be done in a chosen (sensible) frequency range system replaced by an $N$-periodic discrete-time system, with the small (fast) sampling time chosen to be a submultipole $\tau/N$ of the controller sampling time $\tau$. By lifting the $N$-periodic control system, a time-invariant discrete-time transfer function representation is obtained. A similar approach has been used for a controller discretization problem by Keller and Anderson (1992), for which it becomes easy to evaluate the norms.

This technique of fast sampling will allow us to approximate the integrals of (26), taken over one (slow) sampling period $\tau$, by the average of their $N$ sampled values over the period $\tau$ for $N$ sufficiently large. To establish the validity of this procedure, we shall show that these sums converge, as $N \to \infty$, to the integrals (26), using the definition of the Riemann integral. This proof of convergence, in turn, requires that the impulse responses of the operators defined in (27) be continuous and exponentially stable. The following lemma establishes this result. The proof is straightforward and is omitted.

*Lemma* 4.1. Under the hypotheses of Theorem 2.1, and the assumption that $\Omega(\infty) = 0$ and that the closed loops depicted in Figs 1 and 5 are exponentially stable (stability of the closed loop in Fig. 1 implies stability of the closed loop in Fig. 5), the impulse response $h(\cdot, \cdot)$ of the overall system in Fig. 5 is continuous for $j\tau < s < (j+1)\tau$, $t > s$ for every integer $j$, and satisfies $|h(t, s)| \le \alpha \exp[-\beta(t-s)]$ $\forall t > s$, for some positive $\alpha$ and $\beta$.

Now let $h_\alpha(t, s)$ again denote the impulse response of any one of the operators $\partial \mathscr{X}_{k,l}/\partial\alpha$ for $\alpha = a_{ij}$, or $b_{ij}$, or $c_{ij}$, as depicted in Fig. 5. These operators are periodic with period $\tau$. Consider now the system defined in Fig. 6, in which $H_{\tau/N}$ and $\Sigma_{\tau/N}$ are, respectively, a hold operator and a sampler operating at the fast rate $\tau/N$. Thus the system of Fig. 6 has discrete-time inputs spaced $\tau/N$ apart, and a similar output stream. The next lemma expresses its impulse response $h_{d\alpha}(i, j)$ as a function of $h_\alpha(t, s)$ and shows that it is $N$-periodic. Again, the straightforward proof is omitted.

*Lemma* 4.2. Let $h_{d\alpha}(i, j)$ denote the impulse response of the system of Fig. 6, formally $\Sigma_{\tau/N} h_\alpha(t, s) H_{\tau/N}$, where $h_\alpha(t, s)$ is identified with $\partial \mathscr{X}_{k,l}/\partial\alpha$ for same $\alpha$. Then

$$h_{d\alpha}(i, j) = (\tau/N) h_\alpha(i\tau/N, s_{i,j}) \tag{28}$$

for some $s_{i,j} \in (j\tau/N, (j+1)\tau/N)$ and

$$h_{d\alpha}(i+N, j+N) = h_{d\alpha}(i, j). \tag{29}$$

Next, consider a system obtained from that of Fig. 6 by blocking $N$ successive inputs and $N$ successive outputs. Thus, if $u_0, u_1, \ldots$ and $y_0, y_1, \ldots$ denote the scalar input and output sequences of the system of Fig. 6, with rate $N/\tau$ then

$$[u_0 \quad u_1 \quad \cdots \quad u_{N-1}]^T, \quad [u_N \quad u_{N+1} \quad \cdots \quad u_{2N-1}]^T, \quad \ldots$$

and

$$[y_0 \quad y_1 \quad \cdots \quad y_{N-1}]^T, \quad [y_N \quad y_{N+1} \quad \cdots \quad y_{2N-1}]^T, \quad \ldots$$

denote the $N$-vector input and output sequences of the new system, with rate $1/\tau$, $1/N$ times the

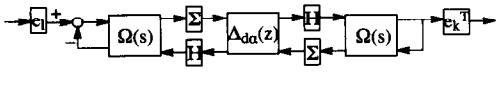


Fig. 5. Redrawing of Fig. 4.

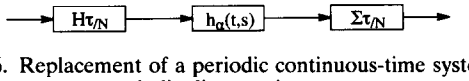


Fig. 6. Replacement of a periodic continuous-time system by a periodic discrete-time system.

rate of the system of Fig. 6. Let $\tilde{h}_\alpha(i,j)$ denote the $N \times N$ impulse response of the new system. A moment's thought shows that

$$[\tilde{h}_\alpha(i,j)]_{p,q} = h_{d\alpha}(Ni + (p-1), Nj + (q-1)). \tag{30}$$

An immediate consequence of the periodicity of $h_{d\alpha}$ established in Lemma 4.2 is the fact that the blocked system is stationary:

$$\tilde{h}_\alpha(i,j) = \tilde{h}_\alpha(i-j, 0).$$

(Henceforth, we shall write $\tilde{h}_\alpha(i-j)$ for $\tilde{h}_\alpha(i-j,0)$.) We remark that, given a state-variable description of the various parts of $h_\alpha(t,s)$, it is easy to get such descriptions for $h_{d\alpha}$ and then $\tilde{h}_\alpha$. This turns out to be important when it comes to evaluating norms.

The above allows us to evaluate $J_B$, at least approximately, using time-invariant quantities. We have the following lemma.

*Lemma* 4.3. With notation as above, the $(i,j)$ entry of

$$J^B_{k,l} = \int_0^\tau \mathrm{d}t \int_{-\infty}^t \frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial B} \left[\frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial B}\right]^{\mathrm{T}} \mathrm{d}s$$

is given by

$$[J^B_{k,l}]_{i,j} = \lim_{N\to\infty} \sum_{m=1}^{L} \sum_{s=0}^{\infty} \mathrm{tr}\,[\tilde{h}_{B\,im}(s)\tilde{h}^{\mathrm{T}}_{B\,jm}(s)],$$

where $\tilde{h}_{B\,im}(s)$ denotes $\tilde{h}_\alpha(s)$ for $\alpha = b_{im}$, the $(i,m)$ entry of $B$.

*Proof.* See the Appendix.

Quantities such as $(J^B_{k,l})_{i,j}$ above are like Gramians, and are comparatively easy to evaluate. Åström and Wittenmark (1990) and Jury (1958) discuss such quantitites, and offer several methods, especially when $i = j$. Let us point out that the identity $\alpha\beta = \frac{1}{4}[(\alpha+\beta)^2 - (\alpha-\beta)^2]$ offers one device to cope with $i \neq j$, if formulae are only available for evaluating sums of squares. Let us also note how simple it is to use linear matrix or Lyapunov equations for evaluation of infinite sums involving a product of two different stable impulse responses. Thus if

$$\alpha_1(k) = h_1^{\mathrm{T}} F_1^{k-1} g_1, \quad \alpha_2(k) = h_2^{\mathrm{T}} F_2^{k-1} g_2$$

then, assuming $|\lambda_i(F_j)| < 1$ for all $i$ and $j = 1, 2$,

$$\sum_{k=1}^{\infty} \alpha_1(k)\alpha_2(k) = h_1^{\mathrm{T}} X h_2,$$

where $X$ solves

$$X - F_1 X F_2^{\mathrm{T}} = g_1 g_2^{\mathrm{T}}.$$

Similarly, it follows that the $(i,j)$ entry of

$$J^C_{k,l} = \int_0^\tau \mathrm{d}t \int_{-\infty}^t \left[\frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial C}\right]^{\mathrm{T}} \frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial C} \mathrm{d}s$$

is given by

$$[J^C_{k,l}]_{i,j} = \lim_{N\to\infty} \sum_{m=1}^{L} \sum_{s=0}^{\infty} \mathrm{tr}\,[\tilde{h}^{\mathrm{T}}_{C\,im}(s)\tilde{h}_{C\,mj}(s)].$$

The calculations are identical to those for $J_B$. It remains to evaluate the first integral of (26).

A typical entry of the matrix

$$\int_0^\tau \mathrm{d}t \int_{-\infty}^t \frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial A} \otimes \frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial A} \mathrm{d}s$$

is given by

$$\int_0^\tau \mathrm{d}t \int_{-\infty}^t h_{A\,mn}(t,s) h_{A\,pq}(t,s)\,\mathrm{d}s,$$

and similar arguments show that this quantity is obtainable as

$$\lim_{N\to\infty} \sum_{s=0}^{\infty} \mathrm{tr}\,[\tilde{h}_{A\,mn}(s)\tilde{h}_{A\,pq}(s)].$$

It remains to explain how to evaluate $\tilde{h}_\alpha$ for $\alpha = a_{ij}$, or $b_{ij}$, or $c_{ij}$. For this purpose, Fig. 7(a–c) is helpful, and illustrates a certain computational simplification. Figure 7(a) is Fig. 6 redrawn, while in Fig. 7(b), which is equivalent to Fig. 7(a), each sampler $\Sigma$ within the hybrid system (which selects a sample every time interval $\tau$) is replaced by a sample $\Sigma_{\tau/N}$, selecting a sample every $\tau/N$ followed by a decimator, which passes through every $N$th input. Each hold of duration $\tau$ is replaced by a repeater (which repeats a signal presented at a given time with the same value $\tau/N, 2\tau/N, \ldots, (N-1)\tau/N$ later) and hold of duration $\tau/N$.

The dashed line encloses a fast discrete-time system (which can be lifted or blocked), and the decimator and repeater serve to connect the discrete-time blocks with different sampling rates.

Lifting produces the arrangement of Fig. 7(c), where the input and output values of $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are obtained by assembling into the one vector $N$ successive values of the input and output of each of the blocks in dashed lines in the set-up of Fig. 7(b).

The listed set-up has $E_1 = [I\ I\ \cdots\ I]^{\mathrm{T}}$ and $E_2 = [I\ 0\ 0\ \cdots\ 0)$ and is a single rate, with period $\tau$, and possesses a transfer function description. The impulse response is $\tilde{h}_\alpha(i)$. Since it is the same set-up as Fig. 7(a) (apart from the

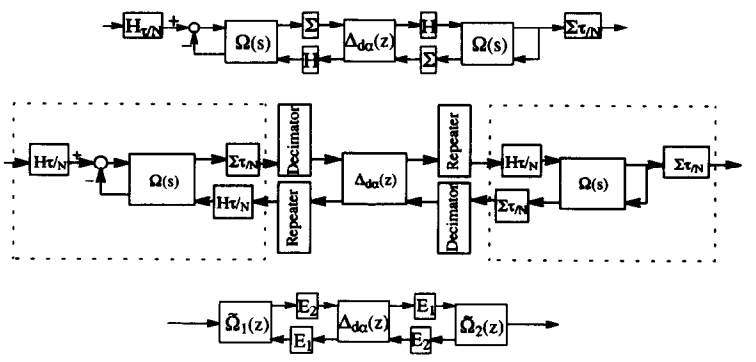


Fig. 7. (a) Redrawing of Fig. 6, (b) replacement of (a) and (c) development of the scheme of Fig. 6 and the lifted system.

way in which inputs and outputs are presented), it is no surprise that norms of equivalent input–output entities are the same. ('Equivalent' means after allowing for the different assembling of inputs/outputs.) The structure of Fig. 7(c) demonstrates that a state-variable realization of $\tilde{h}_\alpha$ comes from assembling realizations of $\tilde{\Omega}_1$, $\tilde{\Omega}_2$ (which are independent of $\alpha$) and $\Delta_{d\alpha}$.

Evidently, the problem of calculating $\partial M_2(P)/\partial P$ is reduced to the problem of calculating infinite sums, which is much simpler, and involves standard computational techniques implemented on most software packages. That makes the iterative algorithm (25) easy to realize.

To summarize, the main steps of the optimal controller realization procedure are as follows, assuming that $K(z)$ is the ideal (infinite-word-length) controller.

*Step* 1. Compute an arbitrary initial realization $(A, B, C, D)$ of $K(z)$ and initialize the iterative algorithm (25) with $P_0 = I$.

*Step* 2. For $k, l = 1, 2, \ldots, L$, approximate $J^B_{k,l}$ of (15) as described in Steps 2.1–2.3, with $N$ sufficiently large and $(J^B_{k,l})_{ij}$ denoting the $(i, j)$ entry of $J^B_{k,l}$. Note that $J^B_{k,l} \in \mathbb{R}^{R\times R}$.

*Step* 2.1. For each $\alpha = b_{im}$, $m = 1, 2, \ldots, L$ and each $\alpha = b_{jm}$, $m = 1, 2, \ldots, L$, let $h_\alpha(t, s)$ be the impulse response defined by Fig. 4. Obtain a state-variable description of $h_\alpha(t, s)$ using state-variable descriptions of the different blocks of Fig. 4. Compute the (fast-sampled) ZOH approximation of this system (see Fig. 6), whose impulse response $h_{d\alpha}(i, j)$ is defined by (26):

$$h_{d\alpha}(i, j) = (\tau/N)h_\alpha(i\tau/N, j\tau/N), \text{ say.}$$

Compute the corresponding $N \times N$ blocked system of Fig. 7(c), whose impulse response is $\tilde{h}_\alpha(i, j) = \tilde{h}(i - j)$, given by (28):

$$[\tilde{h}_\alpha(i - j)]_{p,q} = h_{d\alpha}(Ni + p - 1, Nj + q - 1), \quad p, q = 1, 2, \ldots, N.$$

*Step* 2.2. Approximate $(J^B_{k,l})_{ij}$ by

$$[J^B_{k,l}]_{i,j} \approx \sum_{m=1}^{L} \sum_{s=0}^{\infty} \operatorname{tr}[\tilde{h}_{B\,im}(s)\tilde{h}^{\mathrm{T}}_{B\,jm}(s)],$$

where $\tilde{h}_{B\,im}(s) = \tilde{h}_\alpha(s)$ for $\alpha = b_{im}$, using the Lyapunov equation procedure suggested in Section 4.

*Step* 2.3. Compute $J_B = \sum_{k,l} J^B_{k,l}$.

*Step* 3. Compute (or rather approximate) $J_C$ using a procedure entirely dual to that for $J_B$.

*Step* 4. Compute

$$\int_0^\tau dt \int_{-\infty}^{t} \frac{\partial \mathscr{K}_{k,l}(t, s)}{\partial A} \otimes \frac{\partial \mathscr{K}_{k,l}(t, s)}{\partial A} ds.$$

The approximation of this integral is performed as follows.

*Step* 4.1. For each $\alpha = a_{mn}$, $m, n = 1, 2, \ldots, R$, let $h_\alpha(t, s)$ be the impulse response defined by Fig. 4. By fast sampling and blocking, compute a state-space realization of the discrete-time impulse response of the corresponding fast-sampled and blocked system as in Step 2.2. Denote by $\tilde{h}_{A\,mn}(s)$ the $N \times N$ impulse response matrix of the corresponding blocked system.

*Step* 4.2. The element of

$$\int_0^\tau dt \int_{-\infty}^{t} \frac{\partial \mathscr{K}_{k,l}(t, s)}{\partial A} \otimes \frac{\partial \mathscr{K}_{k,l}(t, s)}{\partial A} ds$$

corresponding to the product of the $(m, n)$ entry of the first matrix with the $(p, q)$ entry of the second matrix is then approximated by

$$\sum_{s=0}^{\infty} \operatorname{tr}[\tilde{h}_{A\,mn}(s)\tilde{h}_{A\,pq}(s)].$$

To compute this infinite sum, the state-space realizations of $\tilde{h}_{A\,mn}(s)$ and $\tilde{h}_{A\,pq}(s)$ are used in combination with the Lyapunov equation technique of Section 4.

*Step* 5. Collect the results of Steps 2–4 in (24) to

compute the gradient $\partial M_2(P)/\partial P$ and update $P_i$ using the iterative algorithm (25).

*Step* 6. Upon convergence of (25) to $P_{opt}$, compute any square root $T_{opt}$ such that $P_{opt} = T_{opt}T_{opt}^T$ and apply the similarity transformation $T_{opt}$ to the initial realization $(A, B, C, D)$ of the compensator $K(z)$ to obtain an optimal realization $(A_{opt}, B_{opt}, C_{opt}, D_{opt})$. Optionally, introduce a further orthogonal transformation to force zero entries into $A_{opt}$, $B_{opt}$ and/or $C_{opt}$ if desired (see e.g. Li *et al.*, 1992).

## 5. NUMERICAL EXAMPLES

We now present two numerical examples to confirm our theoretical results. The first is a simple one with a one-state controller and has no applied interest, but allows us to get better understanding of the system's properties and behaviour. The second example has been used in Ackermann (1985).

### 5.1. *Example* 1.

The plant to be controlled is given by its transfer function

$$\Pi(s) = \frac{s + 0.9531}{s - 0.0953}.$$

The desirable control strategy is to control this plant in such a way that the closed loop has the following transfer function:

$$X(s) = \frac{0.8318}{s + 0.6931}.$$

The controller is to be used with a sampler and a zero-order hold with sampling period $\tau = 1$ has the following transfer function:

$$K(z) = \frac{0.6}{z}.$$

Let us consider two realizations of the controller

$$K(z) = \frac{bc}{z - a} + d.$$

The first is with $a = d = 0$, $b = 0.006$, $c = 100$ and the second is with $a = d = 0$, $b_{opt} = c_{opt} = \sqrt{0.6}$.

The second realization is the optimal one. For this one-state controller both the $\mathcal{L}_2$ measure minimization (using fast sampling and blocking) described in this paper and optimization without fast sampling give the same optimal realization. By optimization without fast sampling, we mean optimization using a discrete-time representation of the plant obtained with the same sampling interval as for the controller. (Equivalently, it is like fast sampling and blocking with $N = 1$!)

When we implement our two realizations of the controller with FWL giving roundoff with two decimal places after the decimal point, we obtain $K(z) = 1/z$ for the first realization and $K_{opt}(z) = 0.59/z$. These controllers give the closed loops

$$X(z) = \frac{1}{z - 0.1}, \quad X_{opt}(z) = \frac{0.59}{z - 0.51}.$$

The frequency responses of these closed loops together with the frequency response of the ideal (realized with infinite precision) closed loop are depicted in Fig. 8. Here, by frequency response of a sampled system we understand the frequency response of the discrete system obtained by interconnecting the discrete controller and a discrete model of the plant, obtained from the zero-order hold equivalent of the continuous model. Though it is not plotted, the closed-loop response of $X(s)$ and $\Pi(s)$ is very close to the 'ideal frequency response' of Fig. 8.

Obviously, the optimal FWL realization gives incomparably better approximation of the desired ideal loop.

### 5.2. *Example* 2. (Ackermann, 1985, p. 239.)

The plant to be controlled is given by its transfer function

$$\Pi(s) = \frac{1}{s + 1}.$$

The controller whose output is the input of a zero-order hold and whose input is sampled with period $\tau = 2$ has the following transfer function:

$$K(z) = \frac{1}{(z - 1)^2}.$$

Consider the realization of the controller given by

$$A = \begin{bmatrix} 4.5105 & -0.7742 \\ 15.918 & -2.5107 \end{bmatrix}, \quad B = \begin{bmatrix} 5.9619 \\ 21.3939 \end{bmatrix},$$

$$C = [-0.8688 \quad 0.2421], \quad D = 0$$

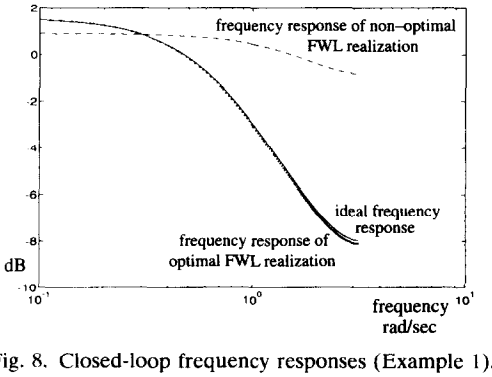


Fig. 8. Closed-loop frequency responses (Example 1).

as an initial one. FWL optimization, not employing fast-sampling/blocking (or, equivalently, employing them for $N = 1$), but using simply a discrete-time representation of the plant and the closed loop, thus neglecting intersample behaviour, gives the following realization:

$$A = \begin{bmatrix} 0.7998 & -0.0404 \\ 0.9981 & 1.1998 \end{bmatrix}, \quad B = \begin{bmatrix} 0.8859 \\ 1.1023 \end{bmatrix},$$

$$C = [-0.9031 \quad 0.7231], \quad D = 0.$$

Minimization of the $M_2$ measure for the pre-fast-sampled and then blocked system gives the optimal realization of the controller given by

$$A = \begin{bmatrix} 0.6556 & -0.2223 \\ 0.5335 & 1.3444 \end{bmatrix}, \quad B = \begin{bmatrix} 1.0507 \\ 0.8551 \end{bmatrix},$$

$$C = [-0.624 \quad 0.7668], \quad D = 0.$$

By using an orthogonal transformation of the state basis, we do not change the formally defined sensitivity and thus optimality. However, by bringing $A$ to Schur form (Li *et al.*, 1992), we can incorporate a zero into the matrix to make computations even more precise, since zero has an infinitely precise computer representation.

The frequency responses of the closed loops corresponding to these two optimal realizations of the closed loop (obtained by different procedures) and the frequency response of the closed loop with the controller given by an initial nonoptimal realization implemented with one decimal place after the decimal point roundoff and the frequency response of the ideal (realized with infinite precision) closed loop are represented in Fig. 9.

The superiority within the passband of the closed-loop system is clearly seen in the optimal sensitivity realization, obtained by fast sampling and blocking of the system, over the optimal sensitivity realization neglecting intersample behaviour of the system. Also, both optimal sensitivity realizations are incomparably superior to the initial nonoptimal realization.

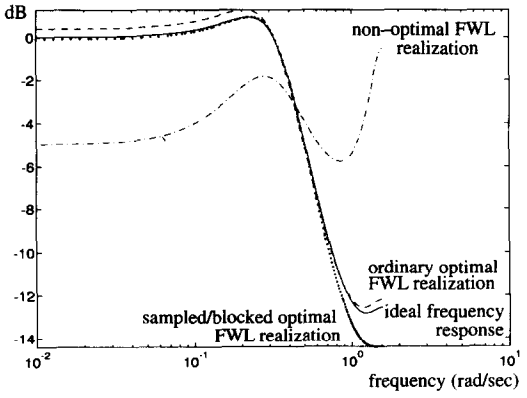


Fig. 9. Closed-loop frequency responses (Example 2).

## 6. CONCLUSIONS

The proposed method obtains the FWL realization of a discrete-time controller, which is used in a closed loop with a continuous-time plant, a sampler, a zero-order hold and an antialiasing filter and which minimizes a sensitivity index. This optimal realization is based on complete information describing the closed-loop system's behaviour, not only at the sampling instances but in intersample periods as well. The existence and uniqueness of this optimal realization (to within an orthogonal coordinate-basis transformation) have been established, and a recursive algorithm converging to the realization has been given.

The theoretical results have been confirmed by two numerical examples, which illustrate the feasibility and efficiency of the proposed method and the advantage of taking into account intersample behaviour of a closed-loop system.

*Acknowledgements*—The authors wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program. This paper presents research results that have been partially supported by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility rests with its authors.

## REFERENCES

Ackermann, J. (1985). *Sampled-Data Control Systems.* Springer, Berlin.

Åström, K. J. and B. Wittenmark (1990). *Computer-Controlled Systems: Theory and Design.* Prentice-Hall, Englewood Cliffs, NJ.

Brewer, J. W. (1978). Kronecker products and matrix calculus in system theory. *IEEE Trans. Circuits Syst.*, **CAS-25,** 772–781.

Francis, B. A. and T. T. Georgiou (1988). Stability theory for linear time-invariant plants with periodic digital controllers. *IEEE Trans. Autom. Control,* **AC-33,** 820–832.

Gevers, M. and G. Li (1993). *Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects.* Springer, London.

Helmke, U. and J. B. Moore (1991). $L^2$ sensitivity minimization of linear system representations via gradient flows. *J. Math. Syst., Estimation and Control,* (to appear).

Hwang, S. Y. (1977). Minimum uncorrelated unit noise in state-space digital filtering. *IEEE Trans. Acoust., Speech, and Signal Processing,* **ASSP-25,** 273–281.

Jury, E. I. (1958). *Sampled-Data Control Systems.* Wiley, New York.

Keller, J. P. and B. D. O. Anderson (1992). A new approach to the discretization of continuous-time controllers. *IEEE Trans. Autom. Control,* **AC-37,** 214–223.

Li, G., B. D. O. Anderson, M. Gevers and J. E. Perkins (1992). Optimal FWL design of state-space digital systems with weighted sensitivity minimization and sparseness consideration, *IEEE Trans. Circuits Syst.*, **CAS-39,** 365–377.

Li, G. and M. Gevers (1990a). Sensitivity and roundoff noise optimization of a state-estimate feedback controller. In *Proc. 11th IFAC World Congress,* Tallinn, **4,** pp. 303–310.

Li, G. and M. Gevers (1990b). Optimal finite precision implementation of a state-estimate feedback controller. *IEEE Trans. Circuits Syst.*, **CAS-38,** 1487–1498.
Li, G. and M. Gevers (1991). Minimization of finite wordlength effects in compensator design. In *Proc. 1st European Control Conference,* Grenoble, pp. 544–549.
Li, G. and M. Gevers (1993). Compensator realizations that minimize the closed loop pole sensitivity. In *Proc. 12th IFAC World Congress,* Sydney, **5,** pp. 149–152.
Liu, K. and R. Skelton (1990). Optimal controllers for finite wordlength implementation. In *Proc. Am. Control Conference,* San Diego.
Liu, K., R. Skelton and K. Grigoriadis (1992). Optimal controllers for finite wordlength implementation. *IEEE Trans. Autom. Control,* **AC-37,** 1294–1304.
Milnor, J. (1963). *Morse Theory.* Princeton University Press, Princeton, NJ.
Mullis, C. T. and R. A. Roberts (1976). Synthesis of minimum roundoff noise fixed-point digital filters. *IEEE Trans. Circuits Syst.*, **CAS-23,** 551–562.
Neudecker, H. (1960). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *J. Am. Statist. Assoc.* **64,** 953–963.
Perkins, J. E., U. Helmke and J. B. Moore (1990). Balanced realizations via gradient flow techniques. *Syst. Control Lett.*, **14,** 369–380.
Thiele, L. (1986). On the sensitivity of linear state-space systems. *IEEE Trans. Circuits Syst.*, **CAS-33,** 502–510.
Williamson, D. (1991). *Digital Control and Implementation: Finite Wordlength Considerations.* Prentice-Hall, Englewood Cliffs, NJ.
Williamson, D. and K. Kadiman (1989). Optimal finite wordlength linear quadratic regulation. *IEEE Trans. Autom. Control,* **AC-34,** 1218–1228.

## APPENDIX—PROOFS

*Proof of Theorem* 3.1

The proof of the theorem can be obtained analogously to a proof of a similar result in Gevers and Li (1993) and Helmke and Moore (1991), but requires certain preliminaries. The first lemma is a variation on a standard result for systems with continuous-time inputs and outputs.

*Lemma A.*1. Consider a periodically time-varying causal linear system $S_1$ with continuous-time input (of arbitrary dimension) and scalar discrete-time output with sampling interval $T$, the underlying period of $S_1$, and a periodically time-varying causal linear system $S_2$ with scalar discrete-time input with the same sampling interval and period, and with continuous-time output (of arbitrary dimension). Denote the impulse response values of $S_1$ and $S_2$ by the row vector $\alpha_1(kT, s)$ and column vector $\alpha_2(t, kT)$, with $k \in \mathbb{Z}$, and $s, t \in \mathbb{R}$. Let $S_3$ denote $S_1$ followed by $S_2$, and have impulse response

$$\alpha_3(t, s) = \sum_{t \geq kT \geq s} \alpha_2(t, kT)\alpha_1(kT, s). \tag{A.1}$$

Then $\alpha_3$ is the zero impulse response if and only if at least one of $\alpha_1$, $\alpha_2$ has this property.

*Proof.* Suppose neither of $\alpha_1$ or $\alpha_2$ is the zero-impulse response. Choose $s$ so that for some $k$, $\alpha_1(kT, s) \neq 0$. Let $k_1$ be the least such $k$. Because $\alpha_2$ is not a zero-impulse response, and is periodically time-varying, $\alpha_2(t, k_1T)$ is not identically zero as a function of $t$. Choose $t \in [(m-1)T, mT]$ for which $\alpha_2(t, k_1T)$ is nonzero and $m$ is minimal. Then

$$\alpha_3(t, s) = \alpha_2(t, k_1T)\alpha_1(k_1T, s) \neq 0. \tag{A.2}$$

This lemma is used solely to establish the next lemma; it will assure us that there is no infimum for $M_2(P)$ involving a singular $P$, or singular $P^{-1}$.

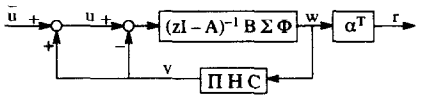


Fig. A.1.

*Lemma A.*2. Under the hypotheses of Theorem 3.1, $J_B$ and $J_C$ are positive definite matrices.

*Proof.* We shall focus on $J_C$ only, the proof for $J_B$ being comparable. By (15) and (17), $J_C$ is nonnegative definite and is singular if and only if for some nonzero control vector $\alpha = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_R]^T$:

$$\frac{\partial \mathcal{X}_{k,l}(t, s)}{\partial C}\alpha = 0 \tag{A.3}$$

for all $t$, $s$, $k$ and $l$. By (10), this means that

$$\mathcal{V}^T e_k e_l^T \mathcal{W}_A^T \alpha = 0 \tag{A.4}$$

for all $t$, $s$, $k$, and $l$. The operator $\mathcal{V}$ is depicted in Fig. 2. It is evident that $\mathcal{V}$ is not identically zero, so that for some choice of $j$ and $k$, $\mathcal{V}_{k,j} = e_k^T \mathcal{V} e_j = e_j^T \mathcal{V}^T e_k$ is not a zero impulse response. If (A.4) holds then left multiplication by $e_l^T$ yields

$$\mathcal{V}_{k,l} e_l^T \mathcal{W}_A^T \alpha = 0, \tag{A.5}$$

and, by Lemma A.1 we must have

$$e_l^T \mathcal{W}_A^T \alpha = 0 \tag{A.6}$$

for all $l$ or

$$\alpha^T \mathcal{W}_A e_l = 0 \tag{A.7}$$

for all $l$. This means that the impulse response of the system $\alpha^T \mathcal{W}_A$ is zero (see Fig. 3), where $\alpha$ is a nonzero constant $R$ vector. The set-up is redrawn in Fig. 10.

For all $u(\cdot)$ in Fig. A.1, $r \equiv 0$. Consider $u(\cdot)$ in Fig. A.1 generated by the arrangement in Fig. A.2. It follows that for all $\bar{u}$ in Fig. A.2, $r \equiv 0$. This is equivalent to having all $\bar{u}$ in Fig. A.3 produce $r \equiv 0$. This can only happen if $\alpha^T(zI - A)^{-1}B = 0$. But since $(A, B)$ is controllable (by the minimality assumption, which is included in the hypothesis of Theorem 3.1), and $\alpha \neq 0$, by assumption above, a contradiction results.

The proof then follows by an application of Theorem 5.1 of Gevers and Li (1993). Alternatively, following ideas of Helmke and Moore (1991), one can obtain a proof as follows. We use the following idea.

*Definition A.*1. Let $J(P)$ be an $n \times n$ matrix $J$. Then the linearization of $J(\cdot)$ at a value of the argument $P$ is that matrix $L$ for which, as $\|\Delta P\| \to 0$

$$\text{vec}\,[J(P + \Delta P) - J(P)] = L \,\text{vec}\,\Delta P + o\,\|\Delta P\|. \tag{A.8}$$

We shall identify $J(P)$ with $\partial M_2(P)/\partial P$ in (24). It is then possible to show that for every positive-definite $P_\infty$ for which $J(P_\infty) = 0$, the linearization $L$ is positive definite, i.e. every extreme point of $M_2(P)$ is a minimum. The properties of $J_B$ and $J_C$ ensure that $M_2(P)$ assumes its global minimum. Morse theory (see e.g. Milnor, 1963), as explained in Helmke and Moore (1991), ensures that there exists a single local minimum, namely the global minimum, of $M_2(P)$. (Consider a smooth real function $m$ of real argument $p$, such that $\lim_{p \to \infty} m(p) = +\infty$ and $\lim_{p \to -\infty} m(p) = +\infty$. If all the extremal points of $m(p)$ are minima then $m(p)$ has a unique minimum. Morse theory is a natural extension of this idea to the multidimensional case.)

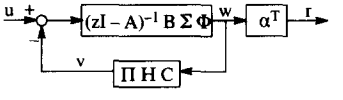


Fig. A.2.

*Proof of Lemma* 4.3

By definition

$$J_B = \int_0^\tau \mathrm{d}t \int_{-\infty}^{t} \frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial B}\left[\frac{\partial \mathscr{K}_{k,l}(t,s)}{\partial B}\right]^{\mathrm{T}} \mathrm{d}s.$$

Hence

$$J_B = \int_0^\tau \mathrm{d}t \int_{-\infty}^{t} \begin{bmatrix} h_{B\,11}(t,s) & \cdots & h_{B\,1L}(t,s) \\ \vdots & & \vdots \\ h_{B\,R1}(t,s) & \cdots & h_{B\,RL}(t,s) \end{bmatrix} \times \begin{bmatrix} h_{B\,11}(t,s) & \cdots & h_{B\,R1}(t,s) \\ \vdots & & \vdots \\ h_{B\,1L}(t,s) & \cdots & h_{B\,RL}(t,s) \end{bmatrix} \mathrm{d}s$$

and

$$[J_B]_{i,j} = \sum_{m=1}^{L} \int_0^\tau \mathrm{d}t \int_{-\infty}^{t} h_{B\,im}(t,s) h_{B\,jm}(t,s)\,\mathrm{d}s.$$

Now let $W$ be a large positive scalar. We shall work temporarily with the following approximation to $(J_B)_{i,j}$:

$$[J_B(W)]_{i,j} = \sum_{m=1}^{L} \int_0^\tau \mathrm{d}t \int_{-W\tau}^{t} h_{B\,im}(t,s) h_{B\,jm}(t,s)\,\mathrm{d}s.$$

The exponent stability of $h$ ensures that $[J_B(W)]_{i,j} \to (J_B)_{i,j}$ as $W \to \infty$. Now

$$[J_B(W)]_{ij} = \sum_{m=1}^{L} \left[ \int_0^\tau \mathrm{d}t \int_0^t h_{B\,im}(t,s) h_{B\,jm}(t,s)\,\mathrm{d}s + \int_0^\tau \mathrm{d}t \int_{-\tau}^{0} h_{B\,im}(t,s) h_{B\,jm}(t,s)\,\mathrm{d}s + \cdots \right.$$

$$\left. + \int_0^t \mathrm{d}t \int_{-W\tau}^{-(W-1)\tau} h_{B\,im}(t,s) h_{B\,jm}(t,s)\,\mathrm{d}s \right].$$


Fig. A.3.

By the continuity properties of $h_\alpha(t,s)$

$$[J_B(W)]_{ij} = \sum_{m=1}^{L} \lim_{N\to\infty} \left[ \frac{\tau^2}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{u-1} h_{B\,im}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) h_{B\,jm}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) + \frac{\tau^2}{N^2} \sum_{u=0}^{N-1} \sum_{v=-N}^{-1} h_{B\,im}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) h_{B\,jm}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) + \cdots + \frac{\tau^2}{N^2} \sum_{u=0}^{N-1} \sum_{v=-Wn}^{-W(N-1)-1} h_{B\,im}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) h_{B\,jm}\left(\frac{u\tau}{N}, \frac{v\tau}{N}\right) \right]$$

$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{v=-Wn}^{u-1} h_{d\,B\,im}(u,v) h_{d\,B\,jm}(u,v)$$

(by Lemma 4.2)

$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} \sum_{s=0}^{W} h_{d\,B\,im}(u, -Ns+2) \times h_{d\,B\,jm}(u, -Ns+t)$$

$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{u=0}^{N-1} \sum_{t=0}^{N-1} \sum_{s=0}^{W} [\tilde{h}_{B\,im}(s,0)]_{ut} \times [\tilde{h}_{B\,jm}(s,0)]_{ut}$$

$$= \sum_{m=1}^{L} \lim_{N\to\infty} \sum_{s=0}^{W} \mathrm{tr}\,[\tilde{h}_{B\,im}(s) \tilde{h}_{B\,jm}^{\mathrm{T}}(s)].$$

The lemma now follows immediately.