International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Stability issues of finite-precision controller structures for sampled-data systems

R. H. Istepanian; S. Chen

Online Publication Date: 15 October 1999
To cite this Article: Istepanian, R. H. and Chen, S. (1999) 'Stability issues of finite-precision controller structures for sampled-data systems', International Journal of Control, 72:15, 1331 - 1342
To link to this article: DOI: 10.1080/002071799220146
URL: http://dx.doi.org/10.1080/002071799220146

INT. J. CONTROL, 1999, VOL. 72, NO. 15, 1331–1342

Stability issues of finite-precision controller structures for sampled-data systems

J. WU†, R. H. ISTEPANIAN‡ and S. CHEN§

The paper investigates the sensitivity of closed-loop stability with respect to (w.r.t.) finite word length (FWL) effects in the implementation of the digital controller coefficients. Both the shift and delta operators are considered for controller parameterization. Two tractable lower-bound measures of closed-loop stability are studied, and the optimal realization of general FWL controller structures is formulated as a constrained non-linear optimization problem. The emphasis of the paper, however, is on the derivation of a new algorithmic approach for the optimal realization of FWL PID controller structures. It is shown that, for PID structures, the optimization can be decoupled into two unconstrained problems with a maximum of four independent variables. An optimization strategy is developed to provide an efficient computational method for searching the optimal FWL PID controller realization with maximum stability bound and minimum bit-length requirement. Simulation results involving an IFAC benchmark PID controller system are presented to illustrate the effectiveness of the proposed strategy.

1. Introduction

The recent advances in fixed-point implementation of digital controllers, such as the design of dedicated fixed-point digital signal processor and digital control processor architectures, have made FWL implementation an important issue in modern control engineering. Improved control performance and increased levels of integration are particularly important in many application areas, such as consumer electronic products, automotive and electromechanical control systems. This is because controller hardware implementation with fixed-point arithmetic offer the advantages of speed, memory space, cost and simplicity over floating-point arithmetic (Masten and Panahi 1997). However, a designed stable closed-loop system may become unstable when the infinite-precision controller is implemented using a fixed-point processor due to FWL effects. The 'robustness' of closed-loop stability w.r.t. controller parameter perturbations therefore is a critical issue in fixed-point implementations and relevant control engineering applications.

In recent years, many results have been reported in the literature dealing with the issues of FWL controller implementation. The degradation effects of FWL on the digital controller designed using an LQG cost function has been investigated (Moroney *et al.* 1980). The effects of FWL implementated digital controller on the stability and performance of sampled-data systems has been analysed (Fialho and Georgiou 1994). A stability measure quantifying the FWL effects has been developed (Moroney *et al.* 1980, Fialho and Georgiou 1994). However, computing explicitly this measure is still an unsolved open problem. To overcome this computational difficulty, two tractable lower bounds of this stability measure have been derived (Li 1998, Istepanian *et al.* 1998 a). The criteria derived provide lower bounds proportional to the closed-loop pole sensitivity measures w.r.t. controller parameter perturbations. It can be shown that the lower bound of Istepanian *et al.* (1998) is a better stability measure than that of Li (1998).

Recent investigations on finite-precision controller realizations have mainly been based on these two lower-bounds of the stability measure and some similar criteria (Madievski *et al.* 1995, Istepanian *et al.* 1996, Li and Gevers 1996, Istepanian *et al.* 1998 b). The present study continues this theme with an emphasis on developing a new optimization method for the optimal realization of finite-precision controller structures. The problem is formulated as a constrained non-linear optimization problem. In particular, for PID sttuctures, the constrained optimization can be decoupled into two unconstrained optimization problems, which permits the development of an effective computational method for obtaining the optimal FWL PID realization with the maximum closed-loop stability measure. Notice that PID controllers have been the most popular controllers in process and industrial control applications for over 50 years and continue to maintain their popularity despite opportunities to apply more advanced control methodologies. This is because of their simplicity, versatility, robustness and successful commercial performance (Åström and Wittenmark 1989). Most of the studies in this area still focus on tuning methods, and very few studies have been reported to date on the FWL

Received 5 September 1997. Revised 11 January 1999. Accepted 23 March 1999.

† National Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou, 310027, P.R. China.

‡ Department of Electrical and Computer Engineering, Ryerson Polytechnic University, 350 Victoria Street, Toronto, Ontario, Canada M5B 2K3.

§ Author for correspondence. Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

International Journal of Control ISSN 0020–7179 print/ISSN 1366–5820 online
http://www.tandf.co.uk/JNLS/con.htm
http://www.taylorandfrancis.com/JNLS/con.htm

implementation issues of discrete PID structures using fixed-point arithmetic (Istepanian 1997).

In all the above-mentioned works addressing the closed-loop stability issues of FWL controller structures, the controllers were described and realized with the usual shift operator. It is known that discrete-time systems can also be described and realized with the delta operator (Middleton and Goodwin 1990). Two major advantages are claimed for the use of δ operator parameterization: a theoretically unified formulation of continuous-time and discrete-time systems; and better numerical implementation properties (Gevers and Li 1993). As with the majority of earlier works, the results presented in this paper, when it was first submitted, were based on the shift operator parameterization. We have since extended the approach to the delta operator parameterization (Wu *et al.* 1999 a,b). These new results are included in this revised paper. Our simulation study confirms that the δ operator parameterization generally results in better closed-loop stability robustness in FWL implementations, compared with the usual shift operator parameterization.

The paper is organized as follows. In § 2, a closed-loop stability measure is presented for sampled-data systems with the shift operator parameterization and FWL implemented controllers. Two tractable lower bounds of this stability measure are considered. The optimal controller realization which maximizes the closed-loop stability measure can be obtained by solving a constrained optimization problem, and this is presented in § 3. Section 4 specifically studies the optimal realization of digital PID controllers subject to FWL constraints. Section 5 extends these results to include the delta operator parameterization. A practical bit length consideration is also discussed. In § 6, the effectiveness of the proposed optimization strategy for PID structures is illustrated by the numerical example of an IFAC benchmark PID control problem (Whidborne *et al.* 1995). Both the shift-operator and delta-operator controllers were tested in the simulation study. Discussions and some concluding remarks are given in § 7.

2. Stability robustness measures of *z* operator based controllers with FWL consideration

Consider the sampled-data system depicted in figure 1, where $P(s)$ is the continuous-time finite-dimensional linear time-invariant plant, $C(z)$ is the discrete-time finite-dimensional linear shift-invariant controller with z indicating the usual shift operator, S_h is the sampler with sampling period h, and H_h is the hold device. The outputs of the sampler and hold device are given by

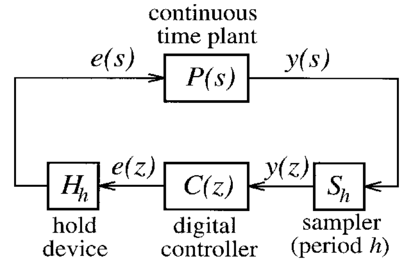

Figure 1. Sampled-data system with digital controller realization.

$$\left.\begin{aligned} y(z) &= S_h y(s): \quad y(k) = y(t)|_{t=kh} \\ e(s) &= H_h e(z): \quad e(t) = e(k) \text{ for } kh < t \leq (k+1)h \end{aligned}\right\} \tag{1}$$

respectively. Assume that $P(s)$ is strictly proper. Let $(A_p, B_p, C_p, 0)$ be a state-space realization of $P(s)$, that is

$$P(s) = C_p(sI - A_p)^{-1}B_p \tag{2}$$

where $A_p \in \mathcal{R}^{m\times m}$, $B_p \in \mathcal{R}^{m\times l}$ and $C_p \in \mathcal{R}^{q\times m}$. Let (A_c, B_c, C_c, D_c) be a state-space realization of $C(z)$, that is

$$C(z) = C_c(zI - A_c)^{-1}B_c + D_c \tag{3}$$

where $A_c \in \mathcal{R}^{n\times n}$, $B_c \in \mathcal{R}^{n\times q}$, $C_c \in \mathcal{R}^{l\times n}$ and $D_c \in \mathcal{R}^{l\times q}$. The state-space realization for a given input–output transfer function is not unique. For example, if (A_c, B_c, C_c, D_c) is a realization of $C(z)$, so is $(\mathcal{T}^{-1}A_c\mathcal{T}, \mathcal{T}^{-1}B_c, C_c\mathcal{T}, D_c)$ for any similarity transformation $\mathcal{T} \in \mathcal{R}^{n\times n}$. Considering the behaviour of the sampled-data system at its sampling instants, we obtain a discrete-time feedback system

$$\left.\begin{aligned} y(z) &= S_h P(s) H_h e(z) \\ e(z) &= C(z)\, y(z) \end{aligned}\right\} \tag{4}$$

The plant $P(z) = S_h P(s) H_h$ is the discretization of $P(s)$, whose state-space realization is $(A_z, B_z, C_z, 0)$ with

$$\left.\begin{aligned} A_z &= e^{A_p h} \in \mathcal{R}^{m\times m} \\ B_z &= \int_0^h e^{A_p \tau} B_p \, d\tau \in \mathcal{R}^{m\times l} \\ C_z &= C_p \in \mathcal{R}^{q\times m} \end{aligned}\right\} \tag{5}$$

It can easily be seen that the corresponding state-space description $(\bar{A}, \bar{B}, \bar{C}, \bar{D})$ of the discrete-time closed-loop system (4) without considering FWL effects is given by

$$\bar{A} = \begin{bmatrix} A_z + B_z D_c C_z & B_z C_c \\ B_c C_z & A_c \end{bmatrix} = \begin{bmatrix} A_z & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} B_z & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} D_c & C_c \\ B_c & A_c \end{bmatrix} \begin{bmatrix} C_z & 0 \\ 0 & I_n \end{bmatrix} = M_0 + M_1 X M_2 = \bar{A}(X) \quad (6)$$

$$\bar{B} = \begin{bmatrix} B_z \\ 0 \end{bmatrix}, \qquad \bar{C} = [C_z \quad 0], \qquad \bar{D} = 0 \quad (7)$$

where $M_0 \in \mathcal{R}^{(m+n)\times(m+n)}$, $M_1 \in \mathcal{R}^{(m+n)\times(l+n)}$ and $M_2 \in \mathcal{R}^{(q+n)\times(m+n)}$ are some fixed matrices that depend on $P(s)$ and h, I_n denotes the $n \times n$ identity matrix, and

$$X = \begin{bmatrix} D_c & C_c \\ B_c & A_c \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & \dots & p_{q+n} \\ p_{q+n+1} & p_{q+n+2} & \dots & p_{2(q+n)} \\ \vdots & \vdots & \dots & \vdots \\ p_{(l+n-1)(q+n)+1} & p_{(l+n-1)(q+n)+2} & \dots & p_{(l+n)(q+n)} \end{bmatrix} \quad (8)$$

will be referred to as the controller matrix.

Suppose that $C(z)$ has been chosen to make the sampled-data system stable and the realization of $C(z)$ is X. Since the sampled-data system is stable if and only if the system (4) is stable (Chen and Francis 1991), it follows that the eigenvalues of $\bar{A}(X)$, denoted by $\{\lambda_i, 1 \le i \le m+n\}$, satisfy

$$|\lambda_i| < 1, \qquad \forall i \in \{1, \dots, m+n\} \quad (9)$$

When the realization (A_c, B_c, C_c, D_c) of $C(z)$ is implemented with a fixed-point processor, the controller matrix X is perturbed into $X + \Delta X$ due to the FWL effects, where

$$\Delta X = \begin{bmatrix} \Delta p_1 & \Delta p_2 & \dots & \Delta p_{q+n} \\ \Delta p_{q+n+1} & \Delta p_{q+n+2} & \dots & \Delta p_{2(q+n)} \\ \vdots & \vdots & \dots & \vdots \\ \Delta p_{(l+n-1)(q+n)+1} & \Delta p_{(l+n-1)(q+n)+2} & \dots & \Delta p_N \end{bmatrix} \quad (10)$$

and $N = (l+n)(q+n)$. Each element of ΔX is bounded, that is

$$\mu(\Delta X) \triangleq \max_{i \in \{1,\dots,N\}} |\Delta p_i| \le \frac{\epsilon}{2} \quad (11)$$

For a fixed-point processor of B_s bits

$$\epsilon = 2^{-(B_s - B_X)} \quad (12)$$

where 2^{B_X} is a 'normalization' factor such that the absolute value of each element of $2^{-B_X} X$ is not larger than 1.

With the perturbation ΔX, λ_i is moved to $\tilde{\lambda}_i$. The closed-loop system is unstable if and only if there exists $i \in \{1, \dots, m+n\}$ such that $|\tilde{\lambda}_i| \ge 1$.

To see when the round-off error will cause the closed-loop system to be unstable, define

$$\mu_0(X) \triangleq \inf\{\mu(\Delta X) : \bar{A}(X) + M_1 \Delta X M_2 \text{ is unstable}\} \quad (13)$$

It quantifies the stability robustness of the realization X to the FWL effects. However, computing explicitly the value of $\mu_0(X)$ is still an unsolved open problem. How 'robust' a controller realization is to the FWL effects can also be viewed from a different angle. Let $B_s^{\min}$ be the smallest word length that can guarantee the closed-loop stability. It would be highly desirable to know $B_s^{\min}$ for a given controller realization. However, except in simulation, it is impractical to test the closed-loop system by reducing B_s until it becomes unstable.

To overcome the difficulty in the computation of $\mu_o(X)$, Istepanian *et al.* (1998 a) introduced a lower bound of $\mu_0(X)$ as

$$\mu_1(X) \triangleq \min_{i \in \{1,\dots,m+n\}} \frac{1 - |\lambda_i|}{\sum_{j=1}^{N} \left| \partial \lambda_i / \partial p_j \big|_X \right|} \quad (14)$$

We have the following theorem.

Theorem 1: *$\bar{A}(X) + M_1 \Delta X M_2$ is stable if $\mu(\Delta X) < \mu_1(X)$.*

Proof: When ΔX is small, using a first-order approximation we have (Li and Gevers 1996, Istepanian *et al.* 1998 a)

$$\Delta \lambda_i = \tilde{\lambda}_i - \lambda_i \approx \sum_{j=1}^{N} \frac{\partial \lambda_i}{\partial p_j}\bigg|_X \Delta p_j, \qquad 1 \le i \le m+n \quad (15)$$

where $\tilde{\lambda}_i$ are the eigenvalues of $\bar{A}(X + \Delta X)$. It follows that

$$|\Delta \lambda_i| \le \sum_{j=1}^{N} \left| \frac{\partial \lambda_i}{\partial p_j}\bigg|_X \right| |\Delta p_j| \le \mu(\Delta X) \sum_{j=1}^{N} \left| \frac{\partial \lambda_i}{\partial p_j}\bigg|_X \right| \quad (16)$$

Thus for $1 \le i \le m+n$, if

$$\mu(\Delta X) < \frac{1 - |\lambda_i|}{\sum_{j=1}^{N} \left| \partial \lambda_i / \partial p_j \big|_X \right|} \quad (17)$$

we have

$$|\tilde{\lambda}_i| \le |\lambda_i| + |\Delta\lambda_i| \le |\lambda_i| + \mu(\Delta X)\sum_{j=1}^{N}\left|\frac{\partial\lambda_i}{\partial p_j}\bigg|_X\right|$$

$$< |\lambda_i| + \frac{1-|\lambda_i|}{\sum_{j=1}^{N}\left|\partial\lambda_i/\partial p_j\big|_X\right|}\sum_{j=1}^{N}\left|\frac{\partial\lambda_i}{\partial p_j}\bigg|_X\right| = 1 \quad (18)$$

which means that $\bar{A}(X+\Delta X)$ is stable. □

The following lemma shows that $\mu_1(X)$ can be computed easily. The proofs of this lemma can be found in Istepanian *et al.* (1998 a).

Lemma 1: *Let $\bar{A}(X)$ be diagonalizable and have $\{\lambda_i, i=1,\ldots,m+n\}$ as its eigenvalues, and $\mathbf{x}_i$ be a right eigenvector of $\bar{A}(X)$ corresponding to the eigenvalue λ_i. Denote $M_x = [\mathbf{x}_1 \cdots \mathbf{x}_{m+n}]$ and $M_y = [\mathbf{y}_1 \cdots \mathbf{y}_{m+n}] = M_x^{-H}$, where $\mathbf{y}_i$ is called the reciprocal left eigenvector corresponding to λ_i, and ^{H} denotes the transpose and conjugate operation. Then $\forall i \in \{1,\ldots,m+n\}$*

$$\frac{\partial\lambda_i}{\partial X} = \begin{bmatrix} \frac{\partial\lambda_i}{\partial p_1} & \frac{\partial\lambda_i}{\partial p_2} & \cdots & \frac{\partial\lambda_i}{\partial p_{q+n}} \\ \frac{\partial\lambda_i}{\partial p_{q+n+1}} & \frac{\partial\lambda_i}{\partial p_{q+n+2}} & \cdots & \frac{\partial\lambda_i}{\partial p_{2(q+n)}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial\lambda_i}{\partial p_{(l+n-1)(q+n)+1}} & \frac{\partial\lambda_i}{\partial p_{(l+n-1)(q+n)+2}} & \cdots & \frac{\partial\lambda_i}{\partial p_N} \end{bmatrix}$$

$$= M_1^T \mathbf{y}_i^* \mathbf{x}_i^T M_2^T \quad (19)$$

*where ^{T} denotes the transpose operation, and * the conjugate operation.*

When a designed infinite-precision stable controller X is implemented with a fixed-point processor, the norm of the controller perturbation $\mu(\Delta X)$ and the lower-bound stability measure $\mu_1(X)$ can be evaluated. If $\mu_1(X) > \mu(\Delta X)$, the closed-loop stability is maintained. Furthermore, when X is implemented with a fixed-point processor of B_s bits, from (11) and Theorem 1, it is easily seen that the closed-loop system is stable if

$$\mu_1(X) > \frac{2^{-(B_s-B_X)}}{2} \quad (20)$$

Define $\hat{B}_{s1}^{\min}$ as the smallest integer that is not less than $-\log_2 \mu_1(X) - 1 + B_X$. We can use $\hat{B}_{s1}^{\min}$ as a super estimate of $B_s^{\min}$. Thus, $\mu_1(X)$ provides a tractable closed-loop stability robustness measure of X with FWL considerations.

Another tractable stability robustness measure with FWL considerations was discussed by Li and Gevers (1996) and Li (1998). This measure is defined as

$$\mu_2(X) \triangleq \min_{i\in\{1,\ldots,m+n\}} \frac{1-|\lambda_i|}{\sqrt{N\sum_{j=1}^{N}\left|\partial\lambda_i/\partial p_j\big|_X\right|^2}} \quad (21)$$

It is also a lower bound of $\mu_0(X)$. Similarly, an estimate $\hat{B}_{s2}^{\min}$ of $B_s^{\min}$ can be computed based on $\mu_2(X)$. Since

$$\left(\sum_{j=1}^{N}\left|\frac{\partial\lambda_i}{\partial p_j}\bigg|_X\right|\right)^2 \le N\sum_{j=1}^{N}\left|\frac{\partial\lambda_i}{\partial p_j}\bigg|_X\right|^2 \quad (22)$$

we have $\mu_2(X) \le \mu_1(X) \le \mu_0(X)$. It is clear that $\mu_1(X)$, which is closer to $\mu_0(X)$, is a better stability robustness measure and can provice a better estimate of $B_2^{\min}$.

3. Optimal realization of *z* operator based controller structures with FWL consideration

From the previous section, we know that there are different realizations X for a given $C(z)$, and the stability measure $\mu_1(X)$ is a function of the realization. It is of practical importance to find a realization such that $\mu_1(X)$ is maximized. Such a realization is optimal in the sense that it has maximum closed-loop stability robustness to FWL effects. The digital controller implemented with an optimal realization means that the stability of the closed-loop system is guaranteed with a minimum hardware requirement in terms of word length. Given an initial realization X_0 of $C(z)$

$$X_0 = \begin{bmatrix} D_c^0 & C_c^0 \\ B_c^0 & A_c^0 \end{bmatrix} \quad (23)$$

any realization of $C(z)$ can be expressed as:

$$X_\mathcal{T} \triangleq \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix} X_0 \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T} \end{bmatrix} \quad (24)$$

where $\mathcal{T} \in \mathcal{R}^{n\times n}$ and $\det(\mathcal{T}) \ne 0$. From (6), the closed-loop transition matrix is

$$\bar{A}(X_\mathcal{T}) = \begin{bmatrix} A_z & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} B_z & 0 \\ 0 & I_n \end{bmatrix}\begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix} \times X_0 \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T} \end{bmatrix}\begin{bmatrix} C_z & 0 \\ 0 & I_n \end{bmatrix}$$

$$= \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix}\begin{bmatrix} A_z & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T} \end{bmatrix} + \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix}\begin{bmatrix} B_z & 0 \\ 0 & I_n \end{bmatrix} \times X_0 \begin{bmatrix} C_z & 0 \\ 0 & I_n \end{bmatrix}\begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T} \end{bmatrix}$$

$$= \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix}\bar{A}(X_0)\begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T} \end{bmatrix} \quad (25)$$

Obviously, $\bar{A}(X_{\mathcal{T}})$ has the same eigenvalues as $\bar{A}(X_0)$. Let λ_i^0 be the ith eigenvalue of $\bar{A}(X_0)$, and $\mathbf{x}_i^0$ and $\mathbf{y}_i^0$ be the corresponding right and reciprocal left eigenvectors, respectively. From (25), the ith right and reciprocal left eigenvectors of $\bar{A}(X_{\mathcal{T}})$ are

$$\begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{-1} \end{bmatrix} \mathbf{x}_i^0 \in \mathcal{C}^{m+n} \quad \text{and} \quad \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \mathbf{y}_i^0 \in \mathcal{C}^{m+n} \tag{26}$$

respectively. Applying Lemma 1, we have

$$\begin{aligned} \left.\frac{\partial \lambda_i}{\partial X}\right|_{X=X_{\mathcal{T}}} &= \begin{bmatrix} B_z^{\mathrm{T}} & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} (\mathbf{y}_i^0)^* (\mathbf{x}_i^0)^{\mathrm{T}} \\ &\quad \times \begin{bmatrix} I_m & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \begin{bmatrix} C_z^{\mathrm{T}} & 0 \\ 0 & I_n \end{bmatrix} \\ &= \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} B_z^{\mathrm{T}} & 0 \\ 0 & I_n \end{bmatrix} (\mathbf{y}_i^0)^* (\mathbf{x}_i^0)^{\mathrm{T}} \\ &\quad \times \begin{bmatrix} C_z^{\mathrm{T}} & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \\ &= \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \left.\frac{\partial \lambda_i}{\partial X}\right|_{X=X_0} \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \end{aligned} \tag{27}$$

From (14), (19) and (27), the optimal realization problem of FWL controllers can be defined as the following maximization problem

$$\varphi \triangleq \max_{X_{\mathcal{T}}} \mu_1(X_{\mathcal{T}}) = \max_{X_{\mathcal{T}}} \min_{1 \le i \le m+n} \frac{1 - |\lambda_i^0|}{\sum_{j=1}^{N} \left| \partial\lambda_i / \partial p_j \big|_{X=X_{\mathcal{T}}} \right|} \tag{28}$$

For the complex-valued matrix $M \in \mathcal{C}^{m \times n}$ with elements $M_{i,j}$, define a norm

$$\|M\|_s \triangleq \sum_{i=1}^{m} \sum_{j=1}^{n} |M_{i,j}| \tag{29}$$

The optimization problem (28) is equivalent to the minimization problem

$$\begin{aligned} \nu = \frac{1}{\varphi} &\triangleq \min_{X_{\mathcal{T}}} \max_{1 \le i \le m+n} \frac{\left\| (\partial\lambda_i/\partial X)|_{X=X_{\mathcal{T}}} \right\|_s}{1 - |\lambda_i^0|} \\ &= \min_{\substack{\mathcal{T} \in \mathcal{R}^{n \times n} \\ \det(\mathcal{T}) \neq 0}} \max_{1 \le i \le m+n} \left\| \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \end{aligned} \tag{30}$$

where

$$\Phi_i = \frac{(\partial\lambda_i/\partial X)|_{X=X_0}}{1 - |\lambda_i^0|}, \qquad 1 \le i \le m+n \tag{31}$$

are the eigenvalue sensitivity matrices.

Thus the optimal controller realization problem is posed as an optimization problem with the cost function

$$f(\mathcal{T}) = \max_{1 \le i \le m+n} \left\| \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \tag{32}$$

and the constraint $\det(\mathcal{T}) \neq 0$. Evidently, the cost function (32) is non-smooth and non-convex, and optimization must be based on a direct search without the aid of cost function derivatives. The conventional optimization methods for this kind of problem, such as Rosenbrock and Simplex algorithms (Kowalik and Osborne 1968, Beveridge and Schechter 1970, Dixon 1972), in general can only find a local minimum. A more serious problem, however, is the need to satisfy $\det(\mathcal{T}) \neq 0$ during the optimization process. It is well known that constrained optimization is more difficult to solve, compared with unconstrained optimization. In all the previous works (Gevers and Li 1993, Li and Gevers 1996, Istepanian *et al.* 1996, 1998 a,b), similar optimization problems were solved with some success by using a conventional direct search algorithm and ignoring the constraints during optimization. Nevertheless, the possible pitfall of violating the constraint by this kind of approach remains, which may result in an invalid solution.

4. Optimal realization of z operator based PID controller structures with FWL consideration

From the previous discussion, we know there is a need to develop some new efficient numerical algorithm for solving the optimal FWL controller realization problem. A main contribution of this paper is to show how this can be achieved for the optimal FWL PID controller realization problem. Assume that $C(z)$ is a digital physically realizable non-interacting PID controller structure (Rad and Lo 1994). Let $(A_c^0 \in \mathcal{R}^{2\times 2}, B_c^0 \in \mathcal{R}^{2\times 1}, C_c^0 \in \mathcal{R}^{1\times 2}, D_c^0 \in \mathcal{R})$ be an initial realization of the PID controller $C(z)$. From (30), the optimal FWL PID controller realization problem is defined as the optimization problem

$$\nu \triangleq \min_{\substack{\mathcal{T} \in \mathcal{R}^{2\times 2} \\ \det(\mathcal{T}) \neq 0}} \max_{1 \le i \le m+2} \left\| \begin{bmatrix} I_l & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} I_q & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \tag{33}$$

As is difficult to handle the constraint $\det(\mathcal{T}) \neq 0$ directly in numerical optimization, we show in the following theorem that the optimization problem (33) can be decoupled into the two 'simpler' unconstrained problems. First we define the two cost functions

$$f_1(x, y, w) = \max_{1 \le i \le m+2} \left\| \begin{bmatrix} w & 0 & 0 \\ 0 & x & 0 \\ 0 & y & 1/x \end{bmatrix} \Phi_i \begin{bmatrix} 1/w & 0 & 0 \\ 0 & 1/x & 0 \\ 0 & -y & x \end{bmatrix} \right\|_s \tag{34}$$

and

$$f_2(x, y, u, w) = \max_{1 \le i \le m+2} \left\| \begin{bmatrix} w & 0 & 0 \\ 0 & x & u \\ 0 & (xy-1)/u & y \end{bmatrix} \times \Phi_i \begin{bmatrix} 1/w & 0 & 0 \\ 0 & y & -u \\ 0 & (1-xy)/u & x \end{bmatrix} \right\|_s \quad (35)$$

Theorem 2: *Let*

$$v_1 = \min_{\substack{x \in (0,+\infty) \\ y \in (-\infty,+\infty) \\ w \in (0,+\infty)}} f_1(x, y, w) \quad (36)$$

and

$$v_2 = \min_{\substack{x \in (-\infty,+\infty) \\ y \in (-\infty,+\infty) \\ u \in (0,+\infty) \\ w \in (0,+\infty)}} f_2(x, y, u, w) \quad (37)$$

Then

$$v = \min\{v_1, v_2\} \quad (38)$$

If $v = v_1$ *and* $(x_{\text{opt1}}, y_{\text{opt1}}, w_{\text{opt1}})$ *is the optimal solution of problem* (36)*, the optimal solution of problem* (33) *is given as*

$$\mathcal{T}_{\text{opt}} = \frac{1}{w_{\text{opt1}}} \begin{bmatrix} x_{\text{opt1}} & y_{\text{opt1}} \\ 0 & 1/x_{\text{opt1}} \end{bmatrix} \quad (39)$$

If $v = v_2$ *and* $(x_{\text{opt2}}, y_{\text{opt2}}, u_{\text{opt2}}, w_{\text{opt2}})$ *is the optimal solution of problem* (37)*, the optimal solution of problem* (33) *is given as*

$$\mathcal{T}_{\text{opt}} = \frac{1}{w_{\text{opt2}}} \begin{bmatrix} x_{\text{opt2}} & (x_{\text{opt2}} y_{\text{opt2}} - 1)/u_{\text{opt2}} \\ u_{\text{opt2}} & y_{\text{opt2}} \end{bmatrix} \quad (40)$$

The proof of Theorem 2 is given in the Appendix. Note that $f_1(x, y, w)$ and $f_2(x, y, u, w)$ are still non-smooth and non-convex functions, and it may be difficult for a conventional non-gradient-based algorithm (Kowalik and Osborne 1968, Beveridge and Schechter 1970, Dixon 1972) to obtain a global minimum solution. This difficulty, however, can be overcome by employing an effcient global optimization strategy, such as the genetic algorithm (GA) (Goldberg 1989, Davis 1991, Man *et al.* 1997) or the adaptive simulated annealing (ASA) (Ingber and Rosen 1992, Ingber 1996, Rosen 1997, Chen *et al.* 1998). In this study, we adopt the ASA for its simplicity and ease of programming. The detailed implementation of the ASA algorithm can be found in Ingber and Rosen (1992), Ingber (1996), Rosen (1997) and Chen *et al.* (1998). It is equally valid to adopt the GA in the optimization.

5. Extension to δ operator parameterization

The results presented in §§ 2–4 are derived based on the z operator parameterization. These have been extended to delta operator based controllers in a new study (Wu *et al.* 1999 a,b). The delta operator is defined as (Middleton and Goodwin 1990)

$$\delta = \frac{z-1}{h} \quad (41)$$

Let a state-space realization of the δ-based controller $C(\delta)$ be $(A_{\delta,c}, B_{\delta,c}, C_{\delta,c}, D_{\delta,c})$. The subscript δ distinguishes this model from the z-based controller realization (A_c, B_c, C_c, D_c). A state-space model of the closed-loop system in the δ domain is $(\bar{A}_\delta, \bar{B}_\delta, \bar{C}_\delta, \bar{D}_\delta)$ with the eigenvalues of $\bar{A}_\delta$ being $\{\lambda_{\delta,i}, 1 \le i \le m+n\}$. Notice that, just as in the z-based case of (6), $\bar{A}_\delta = \bar{A}_\delta(X_\delta)$ is a function of the controller matrix

$$X_\delta = \begin{bmatrix} D_{\delta,c} & C_{\delta,c} \\ B_{\delta,c} & A_{\delta,c} \end{bmatrix} \quad (42)$$

The relationships between the z and δ parameterizations are well known (Neuman 1993 a,b). For example, two state-space models of $C(z)$ and $C(\delta)$ are linked by

$$A_c = hA_{\delta,c} + I_n, \quad B_c = hB_{\delta,c}, \quad C_c = C_{\delta,c}, \quad D_c = D_{\delta,c} \quad (43)$$

and the two sets of the eigenvalues $\{\lambda_i\}$ and $\{\lambda_{\delta,i}\}$ satisfy

$$\lambda_i = 1 + h\lambda_{\delta,i}, \quad \forall i \quad (44)$$

From (44), we have the condition of closed-loop stability in the δ domain.

Lemma 2: *The discrete-time system* $(\bar{A}_\delta, \bar{B}_\delta, \bar{C}_\delta, \bar{D}_\delta)$ *is stable if and only if*

$$\left| \lambda_{\delta,i} + \frac{1}{h} \right| < \frac{1}{h}, \quad \forall i \quad (45)$$

We can now summarize the main results of Wu *et al.* (1999 a,b). Similar to (14), a lower-bound stability measure for δ-based FWL controllers is

$$\mu_1(X_\delta) \triangleq \min_{1 \le i \le m+n} \frac{(1/h) - |\lambda_{\delta,i} + (1/h)|}{\sum_{j=1}^{N} \left| \left. \partial \lambda_{\delta,i} / \partial p_j \right|_{X_\delta} \right|} \quad (46)$$

where p_j are the elements of X_δ. The optimal realization problem of δ-based controller structures with FWL consideration is posed as a constrained optimization problem with the cost function $f(\mathcal{T})$ as defined in (32) and subject to the constraint $\det(\mathcal{T}) \ne 0$, but the eigenvalue sensitivity matrices are now given differently by

$$\Phi_i = \frac{(\partial \lambda_{\delta,i} / \partial X)|_{X = X_{\delta,0}}}{(1/h) - |\lambda_{\delta,i}^0 + (1/h)|} \quad (47)$$

where $X_{\delta,0}$ is the initial realization of the controller matrix and $\lambda^0_{\delta,i}$ the eigenvalues of $\bar{A}_\delta(X_{\delta,0})$. The computation of $\partial\lambda_{\delta,i}/\partial X$ is the same as given in (19) but the matrices M_1 and M_2 are now formed differently from the state-space model of the δ-based plant model $P(\delta)$. It can also be shown that the optimal realization problem of δ-based FWL PID controller structures can be decoupled into two unconstrained optimization problems and a theorem similar to Theorem 2 can be proved (Wu *et al.* 1999 a).

It is worth pointing out a practical constraint on the FWL implementation of δ-based controllers, which is often overlooked. The state-space equation of the δ-based controller

$$\delta x(k) = A_{\delta,c}x(k) + B_{\delta,c}u(k) \quad (48)$$

is realized using

$$x(k+1) = x(k) + h(A_{\delta,c}x(k) + B_{\delta,c}u(k)) \quad (49)$$

The sampling period h should be implemented exactly without any FWL errors in order to avoid further perturbations to the controller X_δ. Otherwise, analysis based on X_δ may not be valid. Notice that controllers based on the z operator do not have this problem.

More specifically, assume that h can be realized exactly by B_h bits with the integer part of h requiring $B_{h,I}$ bits and the fractional part of h requiring $B_{h,F}$ bits. Let $\hat{B}^{\min}_{s1}$ be the smallest integer that is not less than $-\log_2 \mu_1(X_\delta) - 1 + B_X$. Here 2^{B_X} is the normalization factor for X_δ. In the z-based case, we can use $\hat{B}^{\min}_{s1}$ as an estimated minimum bit length that can guarantee the closed-loop stability. In the δ-based case, this needs modification to take into accountt the requirement of implementing$h$ exactly. A modified estimate of the minimum bit length that can guarantee the closed-loop stability is

$$\hat{B}^{\min}_{sh} = \max\{B_{h,I}, B_X\} + \max\{B_{h,F}, \hat{B}^{\min}_{s1} - B_X\} \quad (50)$$

For example, if $h = 2^3$ and $\hat{B}^{\min}_{s1} = 8$ with $B_X = 1$, the estimated minimum bit length is $\hat{B}^{\min}_{sh} = 3 + (8 - 1) = 10$. If $h = 2^{-10}$ and $\hat{B}^{\min}_{s1} = 4$ with $B_X = 1$, $\hat{B}^{\min}_{sh} = 1 + 10 = 11$.

6. A numerical example

To show how the optimization approach presented earlier can be used efficiently for designing optimal FWL PID controller structures, we consider the following IFAC benchmark PID control system (Whidborne *et al.* 1995). The continuous-time plant model is

$$P(s) = \frac{25(-0.4s+1)}{(s^2+3s+25)(5s+1)} \quad (51)$$

and the designed PID controller is

$$C(s) = 1.311 + \frac{0.431}{s} + \frac{1.048s}{1+12.92s} \quad (52)$$

The sampled-data system with the infinite-precision digital controller is stable when the sampling period $h \le 2^3$. The range of the sampling period tested in the simulation was 2^3 to 2^{-12}, to cover the slow to very fast sampling conditions. For the comparison purpose, both the z and δ based controllers were investigated in the simulation. To study the important role of the optimization algorithm employed, both the conventional Rosenbrock and advanced ASA algorithms were used in the optimization.

6.1. *Results for* z *operator based controllers*

Given a sampling rate, the discrete-time plant model $P(z)$ and the digital controller $C(z)$ were obtained. The initial controller realization X_0 was chosen to be the controllable canonical realization. The eigenvalues $\{\lambda_i\}$ of the ideal closed-loop system and the eigenvalue sensitivity matrices $\{\Phi_i\}$ were then computed. The optimal PID controller realizations obtained by solving the optimization problems (36) and (37) with the Rosenbrock algorithm were denoted as $\tilde{X}_{\text{opt1}}$ and $\tilde{X}_{\text{opt2}}$, respectively. Similarly, the two optimal solutions of (36) and (37) obtained using the ASA algorithm were denoted as X_{opt1} and X_{opt2}, respectively. Table 1 summarizes the values of the stability lower bound measure μ_1 for different controller realizations under various sampling conditions, and table 2 lists the corresponding estimated minimum bit lengths that can guarantee the closed-loop stability for these controller realizations.

Several observations can readily be made. The results given in tables 1 and 2 show that the optimal controller realizations have much larger closed-loop stability margins than the non-optimal controllable canonical realization and require much smaller word lengths in fixed-point implementation. In the very fast sampling condition of $h = 2^{-12}$, the stability measure of X_{opt2} is 10^5 times larger than that of X_0. It can also be seen that, when the sampling rate increases, the closed-loop stability measure of the z-based controller decreases considerably. This is true for non-optimal and optimal controller realizations. The ASA algorithm generally yielded better optimization results, compared with the Rosenbrock algorithm. From the results listed in table 1, it is obvious that the conventional Rosenbrock algorithm often missed the true global optimal solution, particularly under fast sampling conditions.

6.2. *Results for* δ *operator-based controllers*

The discrete-time plant model $P(\delta)$ and controller $C(\delta)$ were derived for each sampling rate. The initial

Sampling period h	Stability measure μ_1				
	X_0	$\tilde{X}_{\text{opt1}}$	$\tilde{X}_{\text{opt2}}$	X_{opt1}	X_{opt2}
2^3	1.306 137e-2	3.856 044e-2	3.893 488e-2	3.767 304e-2	3.675 970e-2
2^2	1.738 083e-2	1.128 260e-1	1.244 934e-1	1.495 225e-1	1.641 928e-1
2^1	5.898 659e-3	6.305 634e-2	6.994 177e-2	1.232 973e-1	1.273 720e-1
2^0	1.754 786e-3	2.805 390e-2	3.574 639e-2	5.466 344e-2	7.310 598e-2
2^{-1}	4.819 871e-4	9.198 569e-3	9.684 189e-3	2.580 249e-2	3.771 688e-2
2^{-2}	1.265 127e-4	2.625 682e-3	1.056 738e-2	8.901 379e-3	1.921 549e-2
2^{-3}	3.242 422e-5	1.361 176e-3	8.597 592e-4	6.696 976e-3	9.719 583e-3
2^{-4}	8.208 513e-6	2.757 079e-4	6.607 722e-4	2.931 593e-3	4.889 652e-3
2^{-5}	2.065 125e-6	6.045 046e-5	7.753 055e-4	8.875 924e-4	2.144 777e-3
2^{-6}	5.179 179e-7	1.073 104e-5	1.216 844e-3	5.955 617e-4	1.073 485e-3
2^{-7}	1.296 848e-7	1.512 025e-5	2.101 092e-5	1.523 501e-5	5.331 186e-4
2^{-8}	3.244 692e-8	6.614 090e-7	3.962 129e-6	8.400 932e-6	3.021 479e-4
2^{-9}	8.114 948e-9	1.758 011e-7	7.370 831e-7	5.230 490e-6	1.240 600e-4
2^{-10}	2.029 139e-9	3.318 801e-8	4.552 015e-7	5.204 957e-7	6.892 182e-5
2^{-11}	5.073 338e-10	1.023 806e-8	7.476 508e-9	6.273 487e-8	3.090 558e-5
2^{-12}	1.268 400e-10	2.150 888e-9	2.923 607e-9	1.970 226e-8	1.327 938e-5

Table 1. Lower-bound stability measures of different z operator based controller realizations for various sampling periods.

Sampling period h	Estimated minimum bit length $\hat{B}_{s1}^{\min}$				
	X_0	$\tilde{X}_{\text{opt1}}$	$\tilde{X}_{\text{opt2}}$	X_{opt1}	X_{opt2}
2^3	8	6	6	6	6
2^2	7	5	5	4	4
2^1	8	6	4	4	3
2^0	10	8	5	6	4
2^{-1}	12	10	9	7	5
2^{-2}	13	12	8	8	6
2^{-3}	15	12	13	8	8
2^{-4}	17	16	13	9	8
2^{-5}	19	17	13	11	9
2^{-6}	21	18	10	11	10
2^{-7}	23	18	18	17	11
2^{-8}	25	22	26	18	12
2^{-9}	27	26	23	19	13
2^{-10}	29	28	23	22	14
2^{-11}	31	31	29	26	15
2^{-12}	33	29	31	26	17

Table 2. Estimated minimal bit lengths of different z operator based controller realizations for various sampling periods.

controller realization $X_{\delta,0}$ was chosen to be the 'controllable' or direct-form realization. The eigenvalues $\{\lambda_{\delta,i}\}$ of the ideal closed-loop system and the eigenvalue sensitivity matrices $\{\Phi_i\}$ were then calculated to form the two optimization problems, similar to (36) and (37). The optimal PID controller realizations obtained using the Rosenbrock algorithm were denoted as $\tilde{X}_{\delta,\text{opt1}}$ and $\tilde{X}_{\delta,\text{opt2}}$, respectively, and the two optimal realizations obtained using the ASA algorithm were denoted as $X_{\delta,\text{opt1}}$ and $X_{\delta,\text{opt2}}$, respectively. Table 3 provides the values of the stability lower bound measure for different controller realizations under various sampling conditions, and table 4 gives the corresponding estimated minimum bit lengths that can guarantee the closed-loop stability for these realizations. Both the $\hat{B}_{s1}^{\min}$ and $\hat{B}_{sh}^{\min}$ are listed in table 4.

The results show that the optimal δ-based controller realization has a much better closed-loop stability margin than the non-optimal direct-form realization $X_{\delta,0}$, although the difference is not as great as in the case of z parameterization. The results also confirm that the δ-based controller realizations have better stability bounds than the z-based realizations under fast sampling conditions. Increasing the sampling rate

Sampling period h	Stability measure μ_1				
	$X_{\delta,0}$	$\tilde{X}_{\delta,\text{opt1}}$	$\tilde{X}_{\delta,\text{opt2}}$	$X_{\delta,\text{opt1}}$	$X_{\delta,\text{opt2}}$
2^3	1.477 681e-3	9.853 745e-3	9.990 982e-3	9.852 372e-3	9.986 133e-3
2^2	4.068 193e-3	6.091 837e-2	6.051 241e-2	6.267 552e-2	6.439 696e-2
2^1	5.081 170e-3	6.838 224e-2	6.964 563e-2	6.842 345e-2	7.051 816e-2
2^0	5.721 692e-3	7.119 400e-2	4.360 879e-2	7.136 269e-2	7.310 503e-2
2^{-1}	6.086 598e-3	7.225 940e-2	6.204 462e-2	7.274 529e-2	7.445 603e-2
2^{-2}	6.279 701e-3	7.370 771e-2	6.446 224e-2	7.328 164e-2	7.515 015e-2
2^{-3}	6.379 331e-3	7.175 426e-2	4.686 201e-2	7.353 966e-2	7.549 933e-2
2^{-4}	6.429 949e-3	7.148 362e-2	7.104 903e-2	7.409 947e-2	7.567 885e-2
2^{-5}	6.455 462e-3	7.053 846e-2	6.664 081e-2	7.393 526e-2	7.576 799e-2
2^{-6}	6.468 270e-3	7.404 889e-2	6.680 168e-2	7.396 521e-2	7.581 252e-2
2^{-7}	6.474 687e-3	7.405 083e-2	7.118 904e-2	7.397 951e-2	7.583 418e-2
2^{-8}	6.477 899e-3	7.081 827e-2	6.691 980e-2	7.417 212e-2	7.584 603e-2
2^{-9}	6.479 505e-3	7.407 674e-2	6.685 483e-2	7.411 691e-2	7.585 130e-2
2^{-10}	6.480 309e-3	7.064 170e-2	7.114 443e-2	7.413 070e-2	7.585 433e-2
2^{-11}	6.480 711e-3	7.404 690e-2	7.104 011e-2	7.419 079e-2	7.585 577e-2
2^{-12}	6.480 912e-3	7.064 419e-2	6.683 266e-2	7.422 571e-2	7.585 604e-2

Table 3. Lower-bound stability measures of different δ operator based controller realizations for various sampling periods.

Sampling period h	Estimated minimum bit length $(\hat{B}^{\min}_{s1}, \hat{B}^{\min}_{sh})$				
	$X_{\delta,0}$	$\tilde{X}_{\delta,\text{opt1}}$	$\tilde{X}_{\delta,\text{opt2}}$	$X_{\delta,\text{opt1}}$	$X_{\delta,\text{opt2}}$
2^3	11, 12	8, 9	8, 9	8, 9	8, 9
2^2	9, 9	6, 6	6, 6	5, 5	5, 5
2^1	8, 8	4, 4	4, 4	4, 4	4, 4
2^0	8, 8	5, 5	7, 7	4, 4	4, 4
2^{-1}	8, 8	5, 5	6, 6	4, 4	4, 4
2^{-2}	8, 8	5, 5	5, 5	4, 4	4, 4
2^{-3}	8, 8	4, 4	7, 7	4, 4	4, 4
2^{-4}	8, 8	5, 6	4, 5	4, 5	4, 5
2^{-5}	8, 8	4, 6	5, 7	4, 6	4, 6
2^{-6}	8, 8	5, 8	5, 8	4, 7	4, 7
2^{-7}	8, 8	4, 8	4, 8	4, 8	4, 8
2^{-8}	8, 9	4, 9	5, 10	4, 9	4, 9
2^{-9}	8, 10	4, 10	4, 10	4, 10	4, 10
2^{-10}	8, 11	4, 11	4, 11	4, 11	4, 11
2^{-11}	8, 12	4, 12	4, 12	4, 12	4, 12
2^{-12}	8, 13	4, 13	4, 13	4, 13	4, 13

Table 4. Estimated minimal bit lengths of different δ operator based controller realizations for various sampling periods. $\hat{B}^{\min}_{s1}$ is estimated from μ_1 and the realization, and $\hat{B}^{\min}_{sh}$ is the modified estimate taking into account the implementation of h.

leads to a slightly improved stability margin for the δ-based controller realization. This is in contrast to the z-based controller realization, which has considerably degraded stability margin when the sampling rate increases. Although the estimated minimum bit length based on μ_1 and the realization is consistently 4 for various sampling conditions, the modified estimate of the minimum bit length is larger and increases as the sampling rate increases. Even taking this into account, however, the estimate of the minimum bit length is still smaller than that of the corresponding z-based controller.

7. Conclusions

In this paper we have proposed a new optimal procedure for the sensitivity analysis of closed-loop stability, subject to FWL implemented controller coefficients. It has been shown that the optimal realization of finite-precision digital controllers can be interpreted as a constrained optimization problem. In particular, for finite-precision PID controllers, the optimization can be decoupled into two unconstrained optimization problems. An efficient optimization approach has been developed for solving this optimal FWL PID controller realization problem. The approach is equally valid for the digital controllers based on either z or δ operator parameterization.

The theoretical results have been verified using a numerical example based on an IFAC benchmark PID control system. The results obtained demonstrate that the proposed approach greatly improves the stability robustness with minimum word length characteristics, compared to non-optimal realizations. The important role of an efficient global optimization method in searching for the true optimal controller realization has also been highlighted in the numerical example. The simulation study also confirms that the δ-based controller has clear advantages over the z-based controller in FWL implementation, particularly under fast sampling conditions.

Future work will investigate the extension of the efficient method for obtaining the FWL PID controller realizations presented in this study to FWL higher-order controller realizations. Ongoing work will also explore the integration of the proposed optimization procedure with the closed-loop controller performance and the sparseness consideration of controller realizations. This will provide a multi-objective framework to develop the optimal finite-precision controller realization that possesses the optimal trade-off between minimal computational requirements, improved performance and stability robustness.

Acknowledgments

This work was supported by the UK Royal Society under an ex-agreement China–British research visit Grant (652053.Q606/AJM). The first author, JW, would also like to thank National Natural Science Foundation of China under Grant 69504010 and Cao Guangbiao Foundation of Zhejiang University for their financial supports.

Appendix

Define the $n \times n$ diagonal matrix set

$$\mathcal{S}_n \triangleq \{ U = \mathrm{diag}(u_1, u_2, \ldots, u_n) : u_i \in \{-1, 1\} \; \forall i \in \{1, \ldots, n\} \} \quad (53)$$

From the definition of $\| M \|_s$ (29), we have

Lemma 3: $\forall M \in \mathcal{C}^{m\times n}$, $U_1 \in \mathcal{S}_m$ *and* $U_2 \in \mathcal{S}_n$

$$\| U_1 M \|_s = \| M \|_s \quad \text{and} \quad \| M U_2 \|_s = \| M \|_s \quad (54)$$

Proof of Theorem 2: Define the sets

$$\begin{aligned}
\zeta_0 &\triangleq \left\{ \mathcal{T} = \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} : t_1 \in \mathcal{R}, t_2 \in \mathcal{R}, t_3 \in \mathcal{R}, t_4 \in \mathcal{R}, t_1 t_4 - t_2 t_3 \neq 0 \right\} \\
\zeta_1 &\triangleq \left\{ \mathcal{T} = \begin{bmatrix} t_1 & t_2 \\ 0 & t_4 \end{bmatrix} : t_1 \in \mathcal{R}, t_2 \in \mathcal{R}, t_4 \in \mathcal{R}, t_1 t_4 \neq 0 \right\} \\
\zeta_2 &\triangleq \left\{ \mathcal{T} = \begin{bmatrix} t_1 & t_2 \\ t_3 & t_4 \end{bmatrix} : t_1 \in \mathcal{R}, t_2 \in \mathcal{R}, t_3 \in \mathcal{R}, t_4 \in \mathcal{R}, t_3 \neq 0, t_1 t_4 - t_2 t_3 \neq 0 \right\}
\end{aligned} \quad (55)$$

Construct the optimization problems

$$v_1 \triangleq \min_{\mathcal{T} \in \zeta_1} \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \quad (56)$$

and

$$v_2 \triangleq \min_{\mathcal{T} \in \zeta_2} \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \quad (57)$$

Obviously $\zeta_0 = \zeta_1 \cup \zeta_2$ and, therefore, $v = \min\{v_1, v_2\}$. Define the function $\mathrm{sgn}(\cdot)$

$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (58)$$

Consider the optimization problem (56). Utilizing Lemma 3, $\forall \mathcal{T} \in \zeta_1$ and $\forall i \in \{1, \ldots, m+2\}$ we have

$$\begin{aligned}
&\left\| \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \\
&= \left\| \begin{bmatrix} 1 & 0 & 0 \\ 0 & t_1 & 0 \\ 0 & t_2 & t_4 \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/t_1 & 0 \\ 0 & -t_2/(t_1 t_4) & 1/t_4 \end{bmatrix} \right\|_s \\
&= \left\| \frac{1}{\sqrt{|t_1 t_4|}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mathrm{sgn}(t_1) & 0 \\ 0 & 0 & \mathrm{sgn}(t_4) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & t_1 & 0 \\ 0 & t_2 & t_4 \end{bmatrix} \right. \\
&\quad \times \Phi_i \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/t_1 & 0 \\ 0 & -t_2/(t_1 t_4) & 1/t_4 \end{bmatrix} \\
&\quad \left. \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & \mathrm{sgn}(t_1) & 0 \\ 0 & 0 & \mathrm{sgn}(t_4) \end{bmatrix} \sqrt{|t_1 t_4|} \right\|_s \\
&= \left\| \begin{bmatrix} 1/\sqrt{|t_1 t_4|} & 0 & 0 \\ 0 & \sqrt{|t_1/t_4|} & 0 \\ 0 & \mathrm{sgn}(t_4) t_2/\sqrt{|t_1 t_4|} & \sqrt{|t_4/t_1|} \end{bmatrix} \right. \\
&\quad \left. \times \Phi_i \begin{bmatrix} \sqrt{|t_1 t_4|} & 0 & 0 \\ 0 & \sqrt{|t_4/t_1|} & 0 \\ 0 & -\mathrm{sgn}(t_4) t_2/\sqrt{|t_1 t_4|} & \sqrt{|t_1/t_4|} \end{bmatrix} \right\|_s
\end{aligned} \quad (59)$$

Define

$$\left.\begin{aligned} x &= \sqrt{|t_1|/|t_4|} \in (0, +\infty) \\ y &= \mathrm{sgn}(t_4)\frac{t_2}{\sqrt{|t_1 t_4|}} \in (-\infty, +\infty) \\ w &= \frac{1}{\sqrt{|t_1 t_4|}} \in (0, +\infty) \end{aligned}\right\} \qquad (60)$$

Then

$$f_1(x, y, w) \triangleq \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} w & 0 & 0 \\ 0 & x & 0 \\ 0 & y & 1/x \end{bmatrix} \times \Phi_i \begin{bmatrix} 1/w & 0 & 0 \\ 0 & 1/x & 0 \\ 0 & -y & x \end{bmatrix} \right\|_s = \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s \qquad (61)$$

and

$$v_1 \triangleq \min_{\mathcal{T} \in \zeta_1} \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{\mathrm{T}} \end{bmatrix} \Phi_i \begin{bmatrix} 1 & 0 \\ 0 & \mathcal{T}^{-\mathrm{T}} \end{bmatrix} \right\|_s = \min_{\substack{x \in (0, +\infty) \\ y \in (-\infty, +\infty) \\ w \in (0, +\infty)}} f_1(x, y, w) \qquad (62)$$

If $v = v_1$ and $(x_{\mathrm{opt1}}, y_{\mathrm{opt1}}, w_{\mathrm{opt1}})$ is the solution of the optimization problem (62)

$$v = v_1 = \max_{i \in \{1,\ldots,m+2\}} \left\| \begin{bmatrix} w_{\mathrm{opt1}} & 0 & 0 \\ 0 & x_{\mathrm{opt1}} & 0 \\ 0 & y_{\mathrm{opt1}} & 1/x_{\mathrm{opt1}} \end{bmatrix} \times \Phi_i \begin{bmatrix} 1/w_{\mathrm{opt1}} & 0 & 0 \\ 0 & 1/x_{\mathrm{opt1}} & 0 \\ 0 & -y_{\mathrm{opt1}} & x_{\mathrm{opt1}} \end{bmatrix} \right\|_s = \max_{i \in \{1,\ldots,m+2\}} \left\| \frac{1}{w_{\mathrm{opt1}}} \begin{bmatrix} w_{\mathrm{opt1}} & 0 & 0 \\ 0 & x_{\mathrm{opt1}} & 0 \\ 0 & y_{\mathrm{opt1}} & 1/x_{\mathrm{opt1}} \end{bmatrix} \times \Phi_i \begin{bmatrix} 1/w_{\mathrm{opt1}} & 0 & 0 \\ 0 & 1/x_{\mathrm{opt1}} & 0 \\ 0 & -y_{\mathrm{opt1}} & x_{\mathrm{opt1}} \end{bmatrix} w_{\mathrm{opt1}} \right\|_s \qquad (63)$$

which means that

$$\mathcal{T}_{\mathrm{opt}} = \frac{1}{w_{\mathrm{opt1}}} \begin{bmatrix} x_{\mathrm{opt1}} & y_{\mathrm{opt1}} \\ 0 & 1/x_{\mathrm{opt1}} \end{bmatrix} \qquad (64)$$

is the optimal solution of the problem (33).

By considering (57) in a similar way, we can prove the rest of Theorem 2. □

References

Åström, K. J., and Wittenmark, K., 1989, *Computer Controlled Systems, Theory and Design* (Englewood Cliffs, NJ: Prentice Hall).

Beveridge, G. S. G., and Schechter, R. S., 1970, *Optimization: Theory and Practice* (McGraw-Hill).

Chen, S., Luk, B. L., and Liu, Y., 1998, Application of adaptive simulated annealing to blind channel identification with HOC fitting. *Electronics Letters*, **34**, 234–235.

Chen, T., and Francis, B. A., 1991, Input–output stability of sampled-data systems. *IEEE Transactions on Automatic Control*, **36**, 50–58.

Davis, L. (Ed.), 1991, *Handbook of Genetic Algorithms* (Van Nostrand Reinhold).

Dixon, L. C. W., 1972, *Nonlinear Optimisation* (London: The English Universities Press Ltd).

Fialho, I. J., and Georgiou, T. T., 1994, On stability and performance of sampled-data systems subject to wordlength constraint. *IEEE Transactions on Automatic Control*, **39**, 2476–2481.

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

Goldberg, D. E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Reading, MA.: Addison-Wesley).

Ingber, L., 1996, Adaptive simulated annealing (ASA): lessons learned. *Journal of Control and Cybernetics*, **25**, 33–54.

Ingber, L., and Rosen, B. E., 1992, Genetic algorithms and very fast simulated reannealing: a comparison. *Mathematical and Computer Modelling*, **16**, 87–100.

Istepanian, R. H., 1997, Implementation issues for discrete PID algorithms using shift and delta operators parameterizations. In *Proceedings of the 4th IFAC Workshop Algorithms and Architectures for Real-Time Control* (Vilamoura, Portugal), pp. 117–122.

Istepanian, R. H., Li, G., Wu, J., and Chu, J., 1998 a, Analysis of sensitivity measures of finite-precision digital controller structures with closed-loop stability bounds. *IEE Proceedings of the Control Theory and Applications*, **145**, 472–478.

Istepanian, R. H., Pratt, I., Goodall, R., and Jones, S., 1996, Effect of fixed point parameterization on the performance of active suspension control systems. In *Proceedings of the 13th IFAC World Congress* (San Francisco, USA), pp. 291–295.

Istepanian, R. H., Wu, J., Whidborne, J. F., Yan, J., and Salcudean, S. E., 1998 b, Finite-word-length stability issues of teleoperation motion-scaling control system. In *Proceedings of the UKACC Control'98* (Swansea, UK), pp. 1676–1681.

Kowalik, J., and Osborne, M. R., 1968, *Methods for Unconstrained Optimization Problems* (New York: Elsevier).

Li, G., 1998, On the structure of digital controllers with finite word length consideration. *IEEE Transactions on Automatic Control*, **43**, 689–693.

Li, G., and Gevers, M., 1996, On the structure of digital controllers in sampled data systems with FWL consideration. In *Proceedings of the 35th IEEE Conference on Decision and Control* (Kobe, Japan), pp. 919–920.

Madievski, A. G., Anderson, B. D. O., and Gevers, M., 1995, Optimum realizations of sampled-data controllers for FWL sensitivity minimization. *Automatica*, **31**, 367–379.

Man, K. F., Tang, K. S., Kwong, S., and Halang, W. A., 1997, *Genetic Algorithms for Control and Signal Processing* (London: Springer).

Masten, M. K., and Panahi, I., 1997, Digital signal processors for modern control systems. *Control Engineering Practice*, **5**, 449–458.

Middleton, R. H., and Goodwin, G. C., 1990, *Digital Control and Estimation: A Unified Approach* (New Jersey: Prentice Hall).

Moroney, P., Willsky, A. S., and Houpt, P. K., 1980, The digital implementation of control compensators: the coefficient wordlength issue. *IEEE Transactions on Automatic Control*, **25**, 621–630.

Neuman, C. P., 1993 a, Transformations between delta and forward shift operator transfer function models. *IEEE Transactions on System Man and Cybernetics*, **23**, 295–296.

Neuman, C. P., 1993 b, Properties of the delta operator model of dynamic physical systems. *IEEE Transactions on System Man and Cybernetics*, **23**, 296–301.

Rad, A. B., and Lo, W. L., 1994, Predictive PI controller. *International Journal of Control*, **60**, 953–975.

Rosen, B. E., 1997, Rotationally parameterized very fast simulated reannealing. Submitted to *IEEE Transactions on Neural Networks*, 1997.

Whidborne, J. F., Murad, G., Gu, D. W., and Postlethwaite, I., 1995, Robust control of an unknown plant—the IFAC93 benchmark. *International Journal of Control*, **61**, 589–640.

Wu, J., Istepanian, R. H., Chen, S., Chu, J., and Whidborne, J. F., 1999 a, Optimizing stability bounds of finite-precision controller structures for sampled-data systems in the delta operator domain. Submitted to *IEE Proceedings of the Control Theory and Applications.*

Wu, J., Istepanian, R. H., Chu, J., Whidborne, J. F., Chen, S., and Hu, J., 1999 b, Stability issues of finite precision controller structures using the delta operator for sampled data systems. To be presented at *14th IFAC World Congress* (Beijing, China), July 4–9.