ELSEVIER

Systems & Control Letters 45 (2002) 321–329

SYSTEMS & CONTROL LETTERS

www.elsevier.com/locate/sysconle

# Two approaches based on pole sensitivity and stability radius measures for finite precision digital controller realizations

S. Chen[a,*], J. Wu[b], G. Li[c]

[a] *Department of Electronics and Computer Sciences, University of Southampton, Highfield, Southampton, UK SO17 1BJ*
[b] *National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, People's Republic of China*
[c] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*

Received 10 September 2000; received in revised form 20 August 2001

**Abstract**

This paper compares the two approaches based on pole sensitivity and the complex stability radius measures, respectively, for optimizing the closed-loop stability robustness of digital controllers with respect to finite word length (FWL) errors in fixed-point implementation. Design details and related optimization procedures are derived for the two methods. The two measures, although derived from different motivations, can both be regarded as lower-bound measures of a true but computationally intractable FWL stability measure in some senses. An example is used to verify the theoretical analysis and to illustrate the two designs for determining optimal FWL controller realizations. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords:* Finite word length; Closed-loop stability; Complex stability radius; Pole sensitivity; Optimization

## 1. Introduction

The current controller design methodology often assumes that the controller is implemented exactly, even though in reality a control law can only be realized in finite precision. It is now well-known that a designed stable control system may achieve a lower than predicted performance or even become unstable when the control law is implemented with a finite-precision device. The finite word length (FWL) effect on the closed-loop stability depends on the controller realization structure, and this property can be utilized to select controller realization in order to improve the FWL stability robustness. Currently, two approaches exist for determining the optimal controller realizations in fixed-point implementation, under the criteria of the pole-sensitivity measure [3–6,11,12,16] and the complex stability radius measure [7,8], respectively.

In the first approach, a suitable pole sensitivity measure is used to quantify the FWL effect, leading to a non-linear optimization problem to find an optimal FWL controller realization. Efficient global optimization techniques to solve for this optimization problem are readily available [2–6,15,16]. In the second approach [8], the complex stability radius measure is employed to formulate an optimal FWL controller realization problem that can be represented as a special $H_\infty$ norm minimization problem and solved for with the method of linear matrix inequality (LMI) [1,14].

* Corresponding author. Tel.: +44-23-8059-6660; fax: +44-23-8059-4508.
*E-mail addresses:* sqc@ecs.soton.ac.uk (S. Chen), jwu@iipc.zju.edu.cn (J. Wu), egli@ntu.edu.sg (G. Li).


PII: S0167-6911(01)00189-X

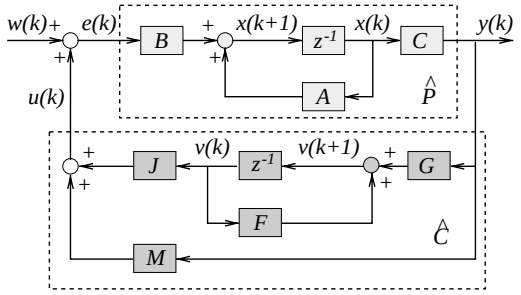


Fig. 1. Discrete-time closed-loop system with a generic output-feedback controller.

This paper provides a comparative study on these two alternative approaches for determining optimal FWL controller realizations.

Detailed design procedures for the two approaches are derived and compared. [1] It can be seen that the pole sensitivity and complex stability radius measures are motivated from different considerations and, in particular, the optimal controller realizations obtained by optimizing the two measures are generally different. However, the both measures involve some approximations in estimating a true stability measure and can therefore be regarded as two "lower-bound" FWL stability measures. Our study shows that the two corresponding optimal controller realizations tend to have similar good FWL characteristics in fixed-point implementation. Advantages and disadvantages of these two alternative methods are discussed and an example is used to illustrate the two design procedures for obtaining optimal FWL controller realizations.

## 2. Problem formulation

Consider the discrete-time closed-loop control system shown in Fig. 1, where the linear time-invariant plant $\hat{P}$ is described by

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{e}(k), \\ \mathbf{y}(k) = \mathbf{C}\mathbf{x}(k), \end{cases} \tag{1}$$

which is completely controllable with $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times p}$ and $\mathbf{C} \in \mathbb{R}^{q\times n}$; and the digital output-feedback controller $\hat{C}$ is described by

$$\begin{cases} \mathbf{v}(k+1) = \mathbf{F}\mathbf{v}(k) + \mathbf{G}\mathbf{y}(k), \\ \mathbf{u}(k) = \mathbf{J}\mathbf{v}(k) + \mathbf{M}\mathbf{y}(k) \end{cases} \tag{2}$$

with $\mathbf{F} \in \mathbb{R}^{m\times m}$, $\mathbf{G} \in \mathbb{R}^{m\times q}$, $\mathbf{J} \in \mathbb{R}^{p\times m}$ and $\mathbf{M} \in \mathbb{R}^{p\times q}$. Assume that a realization $(\mathbf{F}_0, \mathbf{G}_0, \mathbf{J}_0, \mathbf{M}_0)$ of $\hat{C}$ has been designed. It is well-known that the realizations of $\hat{C}$ are not unique. All the realizations of $\hat{C}$ form the set

$$\mathbb{S} = \{(\mathbf{F}, \mathbf{G}, \mathbf{J}, \mathbf{M}) \colon \mathbf{F} = \mathbf{T}^{-1}\mathbf{F}_0\mathbf{T}, \mathbf{G} = \mathbf{T}^{-1}\mathbf{G}_0, \mathbf{J} = \mathbf{J}_0\mathbf{T}, \mathbf{M} = \mathbf{M}_0\}, \tag{3}$$

where $\mathbf{T} \in \mathbb{R}^{m\times m}$ is any real-valued non-singular matrix. Let $\mathbf{w}_\mathbf{F} = \mathrm{Vec}(\mathbf{F})$, with $\mathrm{Vec}(\cdot)$ defining the column stacking operator. Denote

$$\mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{w}_\mathbf{F} \\ \mathbf{w}_\mathbf{G} \\ \mathbf{w}_\mathbf{J} \\ \mathbf{w}_\mathbf{M} \end{bmatrix}, \quad \mathbf{w}_0 \triangleq \begin{bmatrix} \mathbf{w}_{\mathbf{F}_0} \\ \mathbf{w}_{\mathbf{G}_0} \\ \mathbf{w}_{\mathbf{J}_0} \\ \mathbf{w}_{\mathbf{M}_0} \end{bmatrix}, \tag{4}$$

where $N = (m+p)(m+q)$. We also refer to $\mathbf{w}$ as a realization of $\hat{C}$. The stability of the closed-loop system in Fig. 1 depends on the poles of the matrix

$$\bar{\mathbf{A}}(\mathbf{w}) = \begin{bmatrix} \mathbf{A}+\mathbf{B}\mathbf{M}\mathbf{C} & \mathbf{B}\mathbf{J} \\ \mathbf{G}\mathbf{C} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^{-1} \end{bmatrix} \bar{\mathbf{A}}(\mathbf{w}_0) \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{bmatrix}. \tag{5}$$

All the different realizations $\mathbf{w}$ in $\mathbb{S}$ achieve exactly the same set of closed-loop poles if they are implemented with infinite precision. Since the closed-loop system will have been designed to be stable, the eigenvalues

$$|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))| = |\lambda_i(\bar{\mathbf{A}}(\mathbf{w}_0))| < 1 \quad \forall i \in \{1, \ldots, m+n\}. \tag{6}$$

When a $\mathbf{w}$ is implemented with a fixed-point processor, it is perturbed into $\mathbf{w} + \Delta\mathbf{w}$ due to the FWL effect. Each element of $\Delta\mathbf{w}$ is bounded by $\pm\varepsilon/2$,

$$\|\Delta\mathbf{w}\|_\infty \triangleq \max_{i\in\{1,\ldots,N\}} |\Delta w_i| \leqslant \varepsilon/2. \tag{7}$$

For a fixed point processor of $B_s$ bits, let $B_s = B_i + B_f$, where $2^{B_i}$ is a "normalization" factor to make the absolute value of each element of $2^{-B_i}\mathbf{w}$ no larger

[1] Fialho and Georgiou's ACC99 paper [8] only contained the two-page summary. The material for the complex stability radius approach presented at this paper are our interpretation.

than 1. Thus, $B_i$ are bits required for the integer part of a number and $B_f$ are bits used to implement the fractional part of a number. It can easily be seen that

$$\varepsilon = 2^{-B_f}. \tag{8}$$

With the perturbation $\Delta\mathbf{w}$, $\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))$ is moved to $\lambda_i(\bar{\mathbf{A}}(\mathbf{w}+\Delta\mathbf{w}))$. If an eigenvalue of $\bar{\mathbf{A}}(\mathbf{w}+\Delta\mathbf{w})$ is outside the open unit disk, the closed-loop system, designed to be stable, becomes unstable with $B_s$-bit implemented $\mathbf{w}$. It is therefore critical to know when the FWL error will cause closed-loop instability. This ultimately means that we would like to know the largest open "sphere" in the controller perturbation space, within which closed-loop remains stable. The size or radius of this "sphere" is defined by

$$\mu_0(\mathbf{w}) \triangleq \inf\{\|\Delta\mathbf{w}\|_\infty : \bar{\mathbf{A}}(\mathbf{w}+\Delta\mathbf{w}) \text{ is unstable}\}. \tag{9}$$

The larger $\mu_0(\mathbf{w})$ is, the larger FWL error the closed-loop stability can tolerate. Let $B_s^{\min}$ be the smallest word length, when used to implement $\mathbf{w}$, can guarantee the closed-loop stability. An estimate of $B_s^{\min}$ can be obtained as

$$\hat{B}_{s,0}^{\min} = B_i + \mathrm{Int}[-\log_2(\mu_0(\mathbf{w}))] - 1, \tag{10}$$

where the integer $\mathrm{Int}[x] \geqslant x$. It can easily be seen that the closed-loop system remains stable if $\mathbf{w}$ is implemented with a fixed-point processor of at least $\hat{B}_{s,0}^{\min}$. Moreover, $\mu_0(\mathbf{w})$ is a function of the controller realization $\mathbf{w}$, we could search for an optimal realization that maximizes $\mu_0(\mathbf{w})$. However, it is not known how to compute the value of $\mu_0(\mathbf{w})$ given a realization $\mathbf{w}$. A practical solution is to consider a lower bound of the stability measure $\mu_0(\mathbf{w})$ in some sense, which is computationally tractable. This in effect defines a smaller but known stable "sphere" or region in the $\Delta\mathbf{w}$ space. Obviously, the closer such a lower bound is to $\mu_0(\mathbf{w})$, the better. The pole sensitivity and the complex stability radius measures can both be regarded as such lower bounds.

## 3. Pole sensitivity approach

Roughly speaking, how easily the FWL error $\Delta\mathbf{w}$ can cause a stable control system to become unstable is determined by how close $|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|$ are to 1 and how sensitive they are to the controller parameter perturbations. This leads to the following FWL stability measure [6]:

$$\mu_p(\mathbf{w}) \triangleq \min_{i\in\{1,\ldots,m+n\}} \frac{1-|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\alpha_i(\mathbf{w})} \tag{11}$$

with

$$\alpha_i(\mathbf{w}) \triangleq \sum_{\mathbf{X}=\mathbf{F},\mathbf{G},\mathbf{J},\mathbf{M}} \left\| \frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{w}_{\mathbf{X}}} \right\|_1, \tag{12}$$

where, for a vector $\mathbf{x}\in\mathbb{C}^s$, the 1-norm $\|\mathbf{x}\|_1$ is defined as

$$\|\mathbf{x}\|_1 \triangleq \sum_{i=1}^{s} |x_i|. \tag{13}$$

The pole sensitivity measure (11) is an improved version of the measure given in [11], that is, it is less conservative in estimating $\mu_0(\mathbf{w})$.

Define a perturbation subset to the controller realization $\mathbf{w}$

$$\mathbb{P}(\mathbf{w}) \triangleq \{\Delta\mathbf{w}:\ |\lambda_i(\bar{A}(\mathbf{w}+\Delta\mathbf{w}))| - |\lambda_i(\bar{A}(\mathbf{w}))| \leqslant \|\Delta\mathbf{w}\|_\infty \cdot \alpha_i(\mathbf{w})\ \forall i\}. \tag{14}$$

It is straightforward to prove the following proposition.

**Proposition 1.** *$\bar{A}(\mathbf{w}+\Delta\mathbf{w})$ is stable if $\Delta\mathbf{w}\in\mathbb{P}(\mathbf{w})$ and $\|\Delta\mathbf{w}\|_\infty < \mu_p(\mathbf{w})$.*

The requirement for $\Delta\mathbf{w}\in\mathbb{P}(\mathbf{w})$ is not too restricted and $\mathbb{P}(\mathbf{w})$ exists, see the discussions in [5,16]. Defining

$$\rho(\mathbb{P}(\mathbf{w})) \triangleq \inf_{\Delta\mathbf{w}\notin\mathbb{P}(\mathbf{w})} \|\Delta\mathbf{w}\|_\infty, \tag{15}$$

we have the following corollary, the proof of which is straightforward.

**Corollary 1.** *$\mu_p(\mathbf{w}) \leqslant \mu_0(\mathbf{w})$ if $\rho(\mathbb{P}(\mathbf{w})) > \mu_0(\mathbf{w})$.*

It can be seen that $\mu_p(\mathbf{w})$ is a lower bound of $\mu_0(\mathbf{w})$, provided that $\mu_0(\mathbf{w})$ is small enough. The assumption of small $\mu_0(\mathbf{w})$ is generally valid, especially for control systems with fast sampling.

The stability measure $\mu_p(\mathbf{w})$ is computationally tractable, as it can readily be shown that [5]:

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial\mathbf{F}} = [\,\mathbf{0}\quad \mathbf{I}\,]L_i(\mathbf{w})\begin{bmatrix}\mathbf{0}\\ \mathbf{I}\end{bmatrix}, \tag{16}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{G}} = [\,\mathbf{0} \quad \mathbf{I}\,] L_i(\mathbf{w}) \begin{bmatrix} \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} \end{bmatrix}, \tag{17}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{J}} = [\,\mathbf{B}^{\mathrm{T}} \quad \mathbf{0}^{\mathrm{T}}\,] L_i(\mathbf{w}) \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \tag{18}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{M}} = [\,\mathbf{B}^{\mathrm{T}} \quad \mathbf{0}^{\mathrm{T}}\,] L_i(\mathbf{w}) \begin{bmatrix} \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} \end{bmatrix} \tag{19}$$

with

$$L_i(\mathbf{w}) = \frac{\mathrm{Re}[\lambda_i^*(\bar{\mathbf{A}}(\mathbf{w}))\mathbf{y}_i^*(\bar{\mathbf{A}}(\mathbf{w}))\mathbf{x}_i^{\mathrm{T}}(\bar{\mathbf{A}}(\mathbf{w}))]}{|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}, \tag{20}$$

where $\mathbf{x}_i(\bar{\mathbf{A}}(\mathbf{w}))$ and $\mathbf{y}_i(\bar{\mathbf{A}}(\mathbf{w}))$ are the right and reciprocal left eigenvectors related to the $\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))$, respectively, $^*$ denotes the conjugate operation, $^{\mathrm{T}}$ the transpose operator, and $\mathrm{Re}[\cdot]$ the real part. Similar to (10), an estimate of $B_s^{\min}$ can be provided with $\mu_p(\mathbf{w})$ by

$$\hat{B}_{s,p}^{\min} = B_i + \mathrm{Int}[-\log_2(\mu_p(\mathbf{w}))] - 1. \tag{21}$$

Given an initial design $\mathbf{w}_0$, the optimal FWL controller realization that maximizes the stability measure (11) is defined as

$$\mathbf{w}_{\mathrm{opt},p} = \arg\max_{\mathbf{w}\in\mathbb{S}} \mu_p(\mathbf{w}) \tag{22}$$

and the optimization procedure to find a $\mathbf{w}_{\mathrm{opt},p}$ can readily be derived. $\forall i \in \{1,\ldots,m+n\}$, partition

$$\mathbf{x}_i(\bar{\mathbf{A}}(\mathbf{w}_0)) = \begin{bmatrix} \mathbf{x}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)) \\ \mathbf{x}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0)) \end{bmatrix},$$

$$\mathbf{y}_i(\bar{\mathbf{A}}(\mathbf{w}_0)) = \begin{bmatrix} \mathbf{y}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)) \\ \mathbf{y}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0)) \end{bmatrix}, \tag{23}$$

where $\mathbf{x}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)),\ \mathbf{y}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)) \in \mathbb{C}^n$, $\mathbf{x}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0))$, $\mathbf{y}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0)) \in \mathbb{C}^m$. It is easily seen from (5) that

$$\mathbf{x}_i(\bar{\mathbf{A}}(\mathbf{w})) = \begin{bmatrix} \mathbf{x}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)) \\ \mathbf{T}^{-1}\mathbf{x}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0)) \end{bmatrix},$$

$$\mathbf{y}_i(\bar{\mathbf{A}}(\mathbf{w})) = \begin{bmatrix} \mathbf{y}_{i,1}(\bar{\mathbf{A}}(\mathbf{w}_0)) \\ \mathbf{T}^{\mathrm{T}}\mathbf{y}_{i,2}(\bar{\mathbf{A}}(\mathbf{w}_0)) \end{bmatrix}. \tag{24}$$

From (16)–(19), we have

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{F}} = \mathbf{T}^{\mathrm{T}} L_{i,2,2}(\mathbf{w}_0)\mathbf{T}^{-\mathrm{T}}, \tag{25}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{G}} = \mathbf{T}^{\mathrm{T}} L_{i,2,1}(\mathbf{w}_0)\mathbf{C}^{\mathrm{T}}, \tag{26}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{J}} = \mathbf{B}^{\mathrm{T}} L_{i,1,2}(\mathbf{w}_0)\mathbf{T}^{\mathrm{T}}, \tag{27}$$

$$\frac{\partial|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}))|}{\partial \mathbf{M}} = \mathbf{B}^{\mathrm{T}} L_{i,1,1}(\mathbf{w}_0)\mathbf{C}^{\mathrm{T}}, \tag{28}$$

where

$$L_{i,j,l}(\mathbf{w}_0) = \frac{\mathrm{Re}[\lambda_i^*(\bar{\mathbf{A}}(\mathbf{w}_0))\mathbf{y}_{i,j}^*(\bar{\mathbf{A}}(\mathbf{w}_0))\mathbf{x}_{i,l}^{\mathrm{T}}(\bar{\mathbf{A}}(\mathbf{w}_0))]}{|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}_0))|}, \quad j,l = 1,2. \tag{29}$$

Define the following cost function:

$$f(\mathbf{T}) \triangleq \min_{i\in\{1,\ldots,m+n\}} \frac{1-|\lambda_i(\bar{\mathbf{A}}(\mathbf{w}_0))|}{\alpha_i(\mathbf{w})} = \mu_p(\mathbf{w}). \tag{30}$$

The optimal realization problem (22) can then be posed as the following optimization problem:

$$\mathbf{T}_{\mathrm{opt},p} = \arg \max_{\substack{\mathbf{T}\in\mathbb{R}^{m\times m} \\ \det(\mathbf{T})\neq 0}} f(\mathbf{T}). \tag{31}$$

Although $f(\mathbf{T})$ is non-smooth and non-convex, efficient global optimization methods exist for solving for this kind of optimization problem [5,2]. With $\mathbf{T}_{\mathrm{opt},p}$, the optimal realization $\mathbf{w}_{\mathrm{opt},p}$ can readily be calculated.

## 4. Complex stability radius approach

Let $\partial E$ denote the unit circle in the complex plane, and $\bar{\sigma}(\mathbf{U})$ the maximal singular value of the complex-valued matrix $\mathbf{U}$. For a stable matrix $\tilde{\mathbf{A}} \in \mathbb{C}^{(n+m)\times(n+m)}$, i.e. $|\lambda_i(\tilde{\mathbf{A}})| < 1$ for $i = 1,\ldots,n+m$, the complex stability radius of a matrix triple $(\tilde{\mathbf{A}},\tilde{\mathbf{B}},\tilde{\mathbf{C}}) \in \mathbb{C}^{(n+m)\times(n+m)} \times \mathbb{C}^{(n+m)\times(p+m)} \times \mathbb{C}^{(q+m)\times(n+m)}$ is defined as

$$r_{\mathbb{C}}(\tilde{\mathbf{A}},\tilde{\mathbf{B}},\tilde{\mathbf{C}}) = \inf\{\bar{\sigma}(\mathbf{\Delta}):\ \mathbf{\Delta}\in\mathbb{C}^{(p+m)\times(q+m)},\ \tilde{\mathbf{A}} + \tilde{\mathbf{B}}\mathbf{\Delta}\tilde{\mathbf{C}} \text{ is unstable}\}. \tag{32}$$

From [13,9], we have

$$r_{\mathbb{C}}(\tilde{\mathbf{A}},\tilde{\mathbf{B}},\tilde{\mathbf{C}}) = \frac{1}{\sup_{z\in\partial E}\bar{\sigma}(\tilde{\mathbf{C}}(z\mathbf{I}-\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{B}})}. \tag{33}$$

Define the transfer function matrix $\hat{\mathbf{G}} = \tilde{\mathbf{C}}(z\mathbf{I}-\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{B}}$ and the $H_\infty$-norm of $\hat{\mathbf{G}}$ [14] as

$$\|\hat{\mathbf{G}}\|_\infty = \sup_{z\in\partial E}\bar{\sigma}(\tilde{\mathbf{C}}(z\mathbf{I}-\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{B}}). \tag{34}$$

Then,

$$r_{\mathbb{C}}(\tilde{\mathbf{A}},\tilde{\mathbf{B}},\tilde{\mathbf{C}}) = \frac{1}{\|\hat{\mathbf{G}}\|_\infty} \tag{35}$$

and we have the following lemma [14; p. 158].

**Lemma 1.** *Let $\gamma > 0$ be a given scalar. The linear time-invariant discrete-time closed-loop transfer function $\hat{\mathbf{G}}$ satisfies $\|\hat{\mathbf{G}}\|_\infty < \gamma$ if and only if there exists a matrix $\mathbf{X} > 0$ such that*

$$\begin{bmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \gamma^2\mathbf{I}\end{bmatrix} > \begin{bmatrix}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \tilde{\mathbf{C}} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{A}} & \tilde{\mathbf{B}}\\ \tilde{\mathbf{C}} & \mathbf{0}\end{bmatrix}^{\mathrm{T}}. \tag{36}$$

Let $\bar{\mathbf{A}}_0$ be the closed-loop system matrix for an initial controller realization $(\mathbf{F}_0,\mathbf{G}_0,\mathbf{J}_0,\mathbf{M}_0)$. For $(\mathbf{F} = \mathbf{T}^{-1}\mathbf{F}_0\mathbf{T}, \mathbf{G} = \mathbf{T}^{-1}\mathbf{G}_0, \mathbf{J} = \mathbf{J}_0\mathbf{T}, \mathbf{M} = \mathbf{M}_0)$, consider the perturbed controller

$$\begin{bmatrix}\mathbf{M} & \mathbf{J}\\ \mathbf{G} & \mathbf{F}\end{bmatrix} + \mathbf{\Delta}, \tag{37}$$

where the perturbation matrix $\mathbf{\Delta}$ is complex-valued. With (37), the closed-loop system matrix (5) becomes

$$\bar{\mathbf{A}} = \begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^{-1}\end{bmatrix}\bar{\mathbf{A}}_0\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}\end{bmatrix} + \begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}\mathbf{\Delta}\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}, \tag{38}$$

where $\mathbf{I}_s$ denotes the $s\times s$ identity matrix. Denote

$$\tilde{\mathbf{A}}(\mathbf{T}) = \begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^{-1}\end{bmatrix}\bar{\mathbf{A}}_0\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}\end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}, \tag{39}$$

$$\tilde{\mathbf{B}} = \begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix} \in \mathbb{R}^{(n+m)\times(p+m)}, \tag{40}$$

$$\tilde{\mathbf{C}} = \begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix} \in \mathbb{R}^{(q+m)\times(n+m)}, \tag{41}$$

$$\hat{\mathbf{G}}(\mathbf{T}) = \tilde{\mathbf{C}}(z\mathbf{I}-\tilde{\mathbf{A}}(\mathbf{T}))^{-1}\tilde{\mathbf{B}}. \tag{42}$$

Then an alternative optimal FWL realization problem is defined as

$$\max_{\mathbf{T}} r_{\mathbb{C}}(\tilde{\mathbf{A}}(\mathbf{T}),\tilde{\mathbf{B}},\tilde{\mathbf{C}}) = \frac{1}{\min_{\mathbf{T}}\|\hat{\mathbf{G}}(\mathbf{T})\|_\infty} = \frac{1}{\mu}. \tag{43}$$

Consider how to solve for the optimal realization problem (43). From Lemma 1, it can be shown that $\|\hat{\mathbf{G}}(\mathbf{T})\|_\infty < \gamma$ if and only if there exists a positive definite matrix $\mathbf{X} \in \mathbb{R}^{(n+m)\times(n+m)}$ such that

$$\begin{bmatrix}\mathbf{P}_1 & & \\ & \mathbf{I}_q & \\ & & \mathbf{P}_2\end{bmatrix} > \mathbf{M}_\gamma\begin{bmatrix}\mathbf{P}_1 & & \\ & \mathbf{I}_p & \\ & & \mathbf{P}_2\end{bmatrix}\mathbf{M}_\gamma^{\mathrm{T}} \tag{44}$$

subject to

$$\mathbf{P}_1 = \begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}\end{bmatrix}\mathbf{X}\begin{bmatrix}\mathbf{I}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^{\mathrm{T}}\end{bmatrix} > 0 \tag{45}$$

and

$$\mathbf{P}_2 = \mathbf{T}\mathbf{T}^{\mathrm{T}} > 0, \tag{46}$$

where

$$\mathbf{M}_\gamma = \begin{bmatrix}\bar{\mathbf{A}}_0 & \tilde{\mathbf{B}}\\ 1/\gamma\tilde{\mathbf{C}} & \mathbf{0}\end{bmatrix}. \tag{47}$$

The inequality (44) with the constraints $\mathbf{P}_1 > 0$ and $\mathbf{P}_2 > 0$ is an LMI problem [1,14], and numerical algorithms for solving for this kind of problems are readily available, for example, in MATLAB toolbox. Therefore, the optimal value of $\mu$ can be obtained together with the corresponding $\mathbf{P}_{1\,\mathrm{opt}}$ and $\mathbf{P}_{2\,\mathrm{opt}}$ using a bisection search method. This leads to

$$\mathbf{T}_{\mathrm{opt},r} = \mathbf{P}_{2\,\mathrm{opt}}^{1/2} \tag{48}$$

and

$$\mathbf{X}_{\mathrm{opt}} = \begin{bmatrix}\mathbf{I}_n & \\ & \mathbf{T}_{\mathrm{opt},r}^{-1}\end{bmatrix}\mathbf{P}_{1\,\mathrm{opt}}\begin{bmatrix}\mathbf{I}_n & \\ & \mathbf{T}_{\mathrm{opt},r}^{-\mathrm{T}}\end{bmatrix}. \tag{49}$$

With $\mathbf{T}_{\mathrm{opt},r}$, the corresponding optimal controller realization $\mathbf{w}_{\mathrm{opt},r}$ can be obtained.

Unlike the pole-sensitivity measure (11), the complex stability radius measure does not have a direct relationship with the word length, and a statistical word

length was adopted to circumvent this difficulty [7]. It follows from

$$\bar{\sigma}(\mathbf{\Delta}) \leqslant \|\mathbf{\Delta}\|_F, \tag{50}$$

that the closed-loop system is stable if

$$\|\mathbf{\Delta}\|_F < r_{\mathbb{C}}, \tag{51}$$

where $\|\cdot\|_F$ denotes Frobnius-norm. Assume that the elements of $\mathbf{\Delta}$ are independently and uniformly distributed in $[-\varepsilon/2, \varepsilon/2]$. From the central limit theorem, $\|\mathbf{\Delta}\|_F^2$ is approximately normally distributed with mean $E_\Delta = N\varepsilon^2/12$ and variance $D_\Delta^2 = N\varepsilon^4/180$, where $N$ is the number of nonzero random elements in $\mathbf{\Delta}$. Thus

$$\Pr(\|\mathbf{\Delta}\|_F \leqslant Q(\varepsilon)) = 0.9777, \tag{52}$$

where

$$Q(\varepsilon) = \sqrt{E_\Delta + 2D_\Delta} = \varepsilon\sqrt{\frac{N}{12} + \sqrt{\frac{N}{45}}}. \tag{53}$$

The above discussions result in the following proposition:

**Proposition 2.** *The closed-loop system is stable with probability no less than* 0.9777, *provided that the elements of* $\mathbf{\Delta}$ *are bounded absolutely by*

$$\mu_r(\mathbf{w}) = \frac{r_{\mathbb{C}}}{\sqrt{N/3 + 4\sqrt{N/45}}}. \tag{54}$$

Thus, the statistical word length formula using the stability measure (54) leads to the following minimum bit length estimate:

$$\hat{B}_{s,r}^{\min} = B_i + \mathrm{Int}[-\log_2(\mu_r(\mathbf{w}))] - 1. \tag{55}$$

## 5. Comparisons

Both the pole sensitivity and complex stability radius approaches involve some approximations in estimating the true stability measure $\mu_0(\mathbf{w})$. Therefore, $\mu_p(\mathbf{w})$ and $\mu_r(\mathbf{w})$ are conservative measures. As conditions are different for them to be lower bounds of $\mu_0(\mathbf{w})$, it is difficult to say which measure is less conservative in estimating the true minimum bit length. It will generally be case dependent. In particular, the corresponding optimal controller realizations $\mathbf{w}_{\mathrm{opt},p}$ and $\mathbf{w}_{\mathrm{opt},r}$ will generally be different. For the pole sensitivity method, the source of approximation is apparent in the Proof of Corollary 1 (see [5,16]). For the complex stability radius measure, the lower-bound nature of $\mu_r(\mathbf{w})$ is less obvious. In practice, the FWL perturbations are real-valued. Taking $\mathbf{\Delta}$ to be complex-valued will introduce some inaccuracy in estimating the true closed-loop stability robustness of a controller realization.

An important advantage of the complex stability radius measure is that the corresponding optimization problem can be posed as the LMI problem (44), and this LMI problem is easier to solve for than the non-linear optimization problem (31). The latter can have many solutions. The pole sensitivity approach however is applicable to the general controller structure that includes output-feedback and observer-based controllers and that is parameterized either by shift or delta operators [5,3,16,6]. The approach based on the complex stability radius measure at its present form can only be applied to output-feedback controllers, and it is not apparent how to generalize to observer-based controllers or controllers in the delta operator domain. For the controller structure given in Fig. 1, experience shows that the two approaches are often compatible in that the two optimal controller realizations $\mathbf{w}_{\mathrm{opt},p}$ and $\mathbf{w}_{\mathrm{opt},r}$ usually have similarly good FWL characteristics in fixed-point implementation.

## 6. A numerical example

A numerical example was used to compare the two FWL optimal design approaches. The example was a torsional vibration control system given in [10]. Discretizing the continuous-time plant with the sampling period 0.001 yielded the discrete-time plant model:

$$\mathbf{A} = \begin{bmatrix} 0.0 & 0.0 & 1.0 \\ 1.0 & 0.0 & -2.97686 \\ 0.0 & 1.0 & 2.97686 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1.0 \\ 0.0 \\ 0.0 \end{bmatrix},$$

$$\mathbf{C} = [\,0.24863 \quad 0.24621 \quad 0.24143\,]$$

and the initially designed controller was given by

$$\mathbf{F}_0 = \begin{bmatrix} 0.0 & -0.33333 \\ 1.0 & 1.33333 \end{bmatrix}, \quad \mathbf{G}_0 = \begin{bmatrix} 1.0 \\ 0.0 \end{bmatrix},$$

$$\mathbf{J}_0 = [\,-1.20982 \quad -0.41278\,], \quad \mathbf{M}_0 = [1.35120].$$

Table 1
Comparison of the two stability measures, corresponding estimated minimum bit lengths and true minimum bit lengths for the initial and three optimal controller realizations

| Realization | $\mu_p$ | $\hat{B}_{s,p}^{\min}$ | $r_{\mathbb{C}}$ | $\mu_r$ | $\hat{B}_{s,r}^{\min}$ | $B_s^{\min}$ |
|---|---|---|---|---|---|---|
| $\mathbf{w}_0$ | 9.8513e-4 | 10 | 5.3470e-3 | 2.4434e-3 | 9 | **7** |
| $\mathbf{w}_{\text{opt},p1}$ | **8.9321e-3** | **8** | 2.0181e-2 | 9.2219e-3 | 8 | **6** |
| $\mathbf{w}_{\text{opt},p2}$ | **8.9317e-3** | **7** | 2.2827e-2 | 1.0431e-2 | 7 | **4** |
| $\mathbf{w}_{\text{opt},r}$ | 5.0274e-3 | 9 | 2.6305e-2 | **1.2021e-2** | **8** | **6** |

With this initial controller realization $\mathbf{w}_0$ together with the given plant model, the two optimization problems (31) and (44) were formed and solved for. For the pole-sensitivity approach, we give the two typical solutions obtained

$$\mathbf{T}_{\text{opt},p1} = \begin{bmatrix} 4.04705\text{e}+0 & -4.57362\text{e}+0 \\ -1.03464\text{e}+1 & 1.92511\text{e}+1 \end{bmatrix}$$

and

$$\mathbf{T}_{\text{opt},p2} = \begin{bmatrix} -8.18364\text{e}-1 & -3.99776\text{e}+0 \\ 3.44463\text{e}+0 & 1.02204\text{e}+1 \end{bmatrix}.$$

The complex stability radius approach produced the following solution:

$$\mathbf{T}_{\text{opt},r} = \begin{bmatrix} 4.00108\text{e}+0 & -1.90171\text{e}+0 \\ -1.47342\text{e}+1 & -5.16409\text{e}-1 \end{bmatrix}.$$

The corresponding controller realizations $\mathbf{w}_{\text{opt},p1}$, $\mathbf{w}_{\text{opt},p2}$ and $\mathbf{w}_{\text{opt},r}$ are, respectively:

$$\mathbf{F}_{\text{opt},p1} = \begin{bmatrix} 0.71295 & -0.88451 \\ -0.12320 & 0.62038 \end{bmatrix},$$

$$\mathbf{G}_{\text{opt},p1} = \begin{bmatrix} 0.62934 \\ 0.33823 \end{bmatrix},$$

$$\mathbf{J}_{\text{opt},p1} = [\,-0.62540 \quad -2.41321\,],$$

$$\mathbf{M}_{\text{opt},p1} = [1.35120],$$

$$\mathbf{F}_{\text{opt},p2} = \begin{bmatrix} 0.62038 & 0.68013 \\ 0.16022 & 0.71295 \end{bmatrix},$$

$$\mathbf{G}_{\text{opt},p2} = \begin{bmatrix} 1.89030 \\ -0.63710 \end{bmatrix},$$

$$\mathbf{J}_{\text{opt},p2} = [\,-0.43180 \quad 0.61778\,],$$

$$\mathbf{M}_{\text{opt},p2} = [1.35120],$$

$$\mathbf{F}_{\text{opt},r} = \begin{bmatrix} 1.07316 & 0.16668 \\ -0.32475 & 0.26017 \end{bmatrix},$$

$$\mathbf{G}_{\text{opt},r} = \begin{bmatrix} 0.01716 \\ -0.48973 \end{bmatrix},$$

$$\mathbf{J}_{\text{opt},r} = [\,1.24139 \quad 2.51388\,],$$

$$\mathbf{M}_{\text{opt},r} = [1.35120].$$

As expected, the two approaches produced different optimal controller realizations. For the initial and three optimal controller realizations, the true minimal bit lengths $B_s^{\min}$ that can guarantee the closed-loop stability were also determined using a computer simulation method. Table 1 compares the values of the two stability measures $\mu_p$ and $\mu_r$, corresponding estimated minimum bit lengths and true minimum bit lengths for the initial and three optimal controller realizations. It can be seen that $\mathbf{w}_{\text{opt},p1}$ or $\mathbf{w}_{\text{opt},p2}$ is not the optimal solution for the optimization problem based on the complex stability radius measure and, similarly, $\mathbf{w}_{\text{opt},r}$ is not the optimal solution for the optimization problem based on the pole sensitivity measure. The results also show that the two optimization procedures are effective, as $\mathbf{w}_{\text{opt},p1}$, $\mathbf{w}_{\text{opt},p2}$ or $\mathbf{w}_{\text{opt},r}$ have much larger FWL stability margins than the initial design $\mathbf{w}_0$.

We also computed the unit impulse response of the closed-loop control system when the controllers were the infinite-precision implemented $\mathbf{w}_0$ and various FWL implemented realizations. Notice that any realization $\mathbf{w} \in \mathbb{S}$, implemented in infinite precision, will achieve the exact performance of the infinite-precision

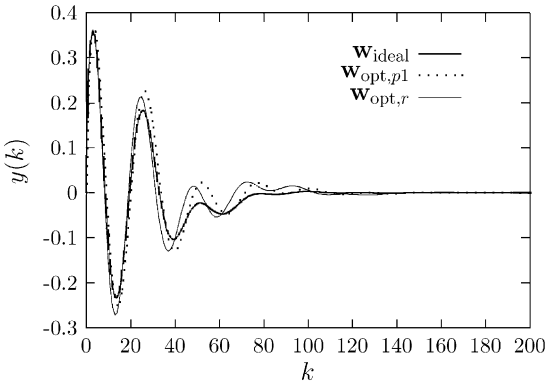


Fig. 2. Comparison of unit impulse response for the infinite-precision controller implementation $\mathbf{w}_{\text{ideal}}$ with those for the 6-bit implemented controller realizations $\mathbf{w}_{\text{opt},p1}$ and $\mathbf{w}_{\text{opt},r}$.

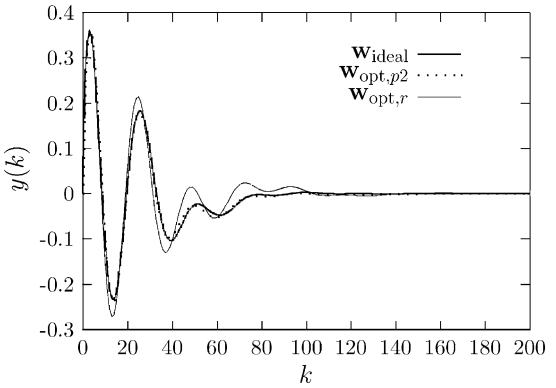


Fig. 3. Comparison of unit impulse response for the infinite-precision controller implementation $\mathbf{w}_{\text{ideal}}$ with those for the 6-bit implemented controller realizations $\mathbf{w}_{\text{opt},p2}$ and $\mathbf{w}_{\text{opt},r}$.

implemented $\mathbf{w}_0$, which is the *designed* controller performance. For this reason, the infinite-precision implemented $\mathbf{w}_0$ is referred to as the *ideal* controller realization $\mathbf{w}_{\text{ideal}}$. Fig. 2 compares the unit impulse response of the plant output for the ideal controller $\mathbf{w}_{\text{ideal}}$ with those for the 6-bit implemented $\mathbf{w}_{\text{opt},p1}$ and $\mathbf{w}_{\text{opt},r}$. The same comparison for $\mathbf{w}_{\text{ideal}}$, $\mathbf{w}_{\text{opt},p2}$ and $\mathbf{w}_{\text{opt},r}$ is given in Fig. 3. Although $\mathbf{w}_{\text{opt},p1}$ and $\mathbf{w}_{\text{opt},r}$ are very different, they both perform similarly well in FWL implementation. It should be emphasized that, although the values of $\mu_r$ for various realizations are consistently larger than those of $\mu_p$, this does not imply that the complex stability radius approach is superior than the pole sensitivity approach. As clearly shown in Fig. 3, $\mathbf{w}_{\text{opt},p2}$ performs better than $\mathbf{w}_{\text{opt},r}$ in a 6-bit implementation.

## 7. Conclusions

In this paper, we have compared the two approaches for obtaining optimal FWL controller realizations based on the pole sensitivity and complex stability radius measures, respectively. Design procedures for the both methods are provided. Although the motivations for these two approaches are different, they can be regarded as two methods of approximating a true FWL closed-loop stability measure. An example is used to compare the two design procedures, and the results show that for the example tested the two approaches produce different optimal controller realizations which have similar good FWL characteristics in fixed-point implementation.

## Acknowledgements

S. Chen and J. Wu wish to thank the support of the UK Royal Society under a KC Wong fellowship (RL/ART/CN/XFI/KCW/11949).

## References

[1] S. Boyd, L. EI Ghaoui, E. Feron, V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, SIAM, Philadelphia, PA, 1994.
[2] S. Chen, B.L. Luk, Adaptive simulated annealing for optimization in signal processing applications, Signal Process. 79 (1) (1999) 117–128.
[3] S. Chen, R.H. Istepanian, J. Wu, J. Chu, Comparative study on optimizing closed-loop stability bounds of finite-precision controller structures with shift and delta operators, Systems Control Lett. 40 (3) (2000) 153–163.
[4] S. Chen, J. Wu, R.H. Istepanian, J. Chu, Optimizing stability bounds of finite-precision PID controller structures, IEEE Trans. Automat. Control 44 (11) (1999) 2149–2153.
[5] S. Chen, J. Wu, R.H. Istepanian, J. Chu, J.F. Whidborne, Optimizing stability bounds of finite-precision controller structures for sampled-data systems in the delta operator domain, IEE Proc. Control Theory Appl. 146 (6) (1999) 517–526.
[6] S. Chen, J. Wu, R.H. Istepanian, G. Li, Optimal finite-precision digital controller realizations based on an improved closed-loop stability measure, in: Proceedings of the UKACC Control 2000, Cambridge, UK, September 4–7, 2000.

[7] I.J. Fialho, T.T. Georgiou, On stability and performance of sampled data systems subject to word length constraint, IEEE Trans. Automat. Control 39 (12) (1994) 2476–2481.
[8] I.J. Fialho, T.T. Georgiou, Optimal finite worldlength digital controller realization, in: Proceedings of the American Control Conference, San Diego, USA, June 2–4, 1999, pp. 4326–4327.
[9] D. Hinrichsen, A.J. Pritchard, Stability radius for structured perturbations and the algebraic Riccati equation, Systems Control Lett. 8 (1986) 105–113.
[10] Y. Hori, A review of torsional vibration control methods and a proposal of disturbance observer-based new techniques, in: Proceedings of the 13th IFAC World Congress, San Francisco, USA, 1996, pp. 7–13.
[11] R.H. Istepanian, G. Li, J. Wu, J. Chu, Analysis of sensitivity measures of finite-precision digital controller structures with closed-loop stability bounds, IEE Proc. Control Theory Appl. 145 (5) (1998) 472–478.
[12] G. Li, On the structure of digital controllers with finite word length consideration, IEEE Trans. Automat. Control 43 (1998) 689–693.
[13] L. Qiu, B. Bernhardsson, A. Rantzer, E.J. Davison, P.M. Young, J.C. Doyle, On the real structured stability radius, in: Proceedings of the 12th IFAC World Congress, Sydney, Australia, 1993, Vol. 8, pp. 71–78.
[14] R.E. Skelton, T. Iwasaki, K.M. Grigoriadid, A Unified Algebraic Approach to Linear Control Design, Taylor and Francis, London, 1998.
[15] J.F. Whidborne, A genetic algorithm approach to designing finite-precision PID controller structures, in: Proceedings of the American Control Conference, San Diego, USA, June 2–4, 1999, pp. 4338–4342.
[16] J. Wu, S. Chen, G. Li, J. Chu, Optimal finite-precision state-estimate feedback controller realization of discrete-time systems, IEEE Trans. Automat. Control 45 (8) (2000) 1550–1554.