Proceedings of the 6th World Congress on Control
and Automation, June 21 - 23, 2006, Dalian, China

Solving Finite Word Length Realization Problems in the Framework of Structured Singular Value

Jun Wu and Jian Chu
National Lab. Ind. Contr. Tech.
Ins. Adv. Process Contr.
Zhejiang Univ., Hangzhou, China
{jwu & chuj}@iipc.zju.edu.cn

Gang Li
School Electri. & Elec. Eng.
Nanyang Tech. Univ.
Singapore
egli@ntu.edu.sg

Sheng Chen
School Elec. & Computer Sci.
University of Southampton
Highfield, Southampton SO17 1BJ, U.K.
sqc@ecs.soton.ac.uk

Abstract - **Based on the structured singular value theory, a tractable stability measure is derived for controllers/filters with the finite word length implementation consideration. The optimal realizations of controllers/filters are defined as those that maximize this measure. A sophisticated optimization strategy is presented to provide an efficient method for solving this problem based on the linear matrix inequality, and a numerical example is given to illustrate the design procedure.**

Index Terms - Finite word length, digital control, controller realization, structured singular value

I. Introduction

It is well-known that the finite word length (FWL) effects cannot be ignored in digital system designs [1],[2]. For example, Keel and Bhattacharyya [3] showed that the digital controller obtained by robust control theory exhibits a poor stability margin with respect to the controller coefficients, if the design does not take into account the FWL implementation related uncertainty properly. A filter/control law can be implemented with different realizations, and these realizations are equivalent if they are implemented in infinite precision. However, different realizations possess different degrees of stability robustness to FWL errors. An FWL design is to select optimal realizations for the given filter/control law by optimizing some FWL stability measures, such as the Frobenius-norm pole sensitivity measure υ_f [4], the l_1-based stability measure υ_l [5], the 1-norm pole sensitivity measure υ_1 [6],[7], the stability radius measure υ_r [8] and the pole sensitivity sum measure υ_s [9]. In fact, the FWL stability measure υ proposed in [10] quantifies the FWL stability characteristics of a realization best. Unfortunately, except for few special cases, how to calculate the value of υ for a given realization is unknown.

Since the computation of the true FWL stability measure υ is an open problem, various tractable FWL stability measures mentioned above are adopted in practice to replace υ. The measures υ_f, υ_1 and υ_s estimate υ through local linearizations of the nonlinear relationship between the system matrix coefficients and system poles, and hence these measures may not always guarantee to be lower bounds of υ. In other words, the minimum word length estimated from υ_f, υ_1 or υ_s may not always maintain stability. The measure υ_r is not surely a lower bound of υ either, because υ_r only provides a statistical word length guaranteeing stability with probability no less than 0.9777. The measure υ_l based on l_1 theory [11] is a lower bound of υ. However, due to the lack of efficient computational tool for l_1 theory, costly numerical methods have to be used to solve the non-convex problem of maximizing υ_l in order to obtain an optimal realization. Structured singular value (SSV) analysis [12],[13] is an important approach of studying stability robustness and linear matrix inequality (LMI) techniques are powerful computational tools for SSV analysis. We propose an SSV-based FWL stability measure υ_μ, which is guaranteed to be a lower bound of υ. The optimal realization problem of optimizing υ_μ can be easily solved using LMI toolboxes of MATLAB. A numerical example is given to illustrate the proposed design method.

II. Notations and preliminaries

Let $\mathcal{R}$ denote the field of real numbers, $\mathcal{C}$ the field of complex numbers, $\mathbf{M}^{\mathrm{T}}$ the transpose of $\mathbf{M}$, $\mathbf{M}^*$ the complex conjugate transpose of $\mathbf{M}$, and $\|\mathbf{M}\|_{\mathrm{m}}$ the maximum absolute value of all the elements in $\mathbf{M}$. Let $\overline{\sigma}$ represent the largest singular value of a matrix, and ρ the spectral radius of a matrix. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, while $\mathbf{I}$ and $\mathbf{0}$ represent the identity and zero matrices of proper dimensions, respectively. $\sharp$ within a matrix represents the symmetric term of the matrix. A discrete-time system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E})$ (or the matrix $\mathbf{A}$) is said stable if $\rho(\mathbf{A}) < 1$. The H_∞-norm of this system is defined as

$$\|\mathbf{E}+\mathbf{C}(z\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}\|_\infty \triangleq \sup_{\substack{z\in\mathcal{C}\\ |z|\geq 1}} \overline{\sigma}[\mathbf{E}+\mathbf{C}(z\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}]. \quad (1)$$

Lemma 1: For stable $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{E})$ with $\|\mathbf{E} + \mathbf{C}(z\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\|_\infty < 1$, there exists a $\mathbf{P} = \mathbf{P}^{\mathrm{T}} > 0$ such that

$$\begin{bmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} - \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{E}\end{bmatrix}^{\mathrm{T}} \begin{bmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{E}\end{bmatrix} > 0. \quad (2)$$

Lemma 2: A real symmetric matrix is partitioned as $\begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{12}^{\mathrm{T}} & \mathbf{A}_{22}\end{bmatrix}$ where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square. This matrix

is positive definite if and only if $\mathbf{A}_{22}$ is positive definite and $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^{\mathrm{T}} > 0$.

The following results of SSV is from [13]. Suppose that we have a matrix $\mathbf{M} \in \mathcal{C}^{n\times n}$ and two non-negative integers p and q with $p+q \leq n$, which specify the number of uncertainty blocks of each type. Then the block structure $\mathbf{k}(p,q)$ is a $p+q$-tuple of positive integers $\mathbf{k} \triangleq [\,k_1 \quad \cdots \quad k_p \quad k_{p+1} \quad \cdots \quad k_{p+q}\,]^{\mathrm{T}}$. This $p+q$-tuple specifies the dimensions of the perturbation blocks, and we require $\sum_{i=1}^{p+q} k_i = n$ in order that these dimensions are compatible with $\mathbf{M}$. Define

$$\mathbf{\Upsilon} = \mathrm{diag}\{\zeta_1\mathbf{I}_{k_1}, \cdots, \zeta_p\mathbf{I}_{k_p}, \xi_{p+1}\mathbf{I}_{k_{p+1}}, \cdots, \xi_{p+q}\mathbf{I}_{k_{p+q}}\} \quad (3)$$

and $\mathcal{K} \triangleq \{\mathbf{\Upsilon} : \zeta_i \in \mathcal{C}, \xi_j \in \mathcal{R}\}$. This determines the set of allowable perturbations. The SSV, $\mu_{\mathbf{k}}(\mathbf{M})$, of $\mathbf{M} \in \mathcal{C}^{n\times n}$ with respect to a block structure $\mathbf{k}(p,q)$ is defined as

$$\mu_{\mathbf{k}}(\mathbf{M}) \triangleq \left(\inf_{\mathbf{\Upsilon}\in\mathcal{K}} \{\overline{\sigma}(\mathbf{\Upsilon}) : \det(\mathbf{I} - \mathbf{\Upsilon}\mathbf{M}) = 0\}\right)^{-1} \quad (4)$$

with $\mu_{\mathbf{k}}(\mathbf{M}) = 0$ if no $\mathbf{\Upsilon} \in \mathcal{K}$ solves $\det(\mathbf{I} - \mathbf{\Upsilon}\mathbf{M}) = 0$.

Lemma 3: If $p = 1$ and $q = 0$, then $\mu_{\mathbf{k}}(\mathbf{M}) = \overline{\sigma}(\mathbf{M})$.

Except for few special cases, the computation of $\mu_{\mathbf{k}}(\mathbf{M})$ is still an open problem. However, an upper bound of $\mu_{\mathbf{k}}(\mathbf{M})$ has been provided, which is easy to compute. Define

$$\mathcal{D}_{\mathbf{k}} \triangleq \mathrm{diag}\{\mathbf{D}_1, \cdots, \mathbf{D}_p, \mathbf{D}_{p+1}, \cdots, \mathbf{D}_{p+q}\} \quad (5)$$

where $0 < \mathbf{D}_i \in \mathcal{C}^{k_i\times k_i}$ and $\mathcal{D}_{\mathbf{k}} \in \mathcal{C}^{n\times n}$,

$$\mathcal{G}_{\mathbf{k}} \triangleq \mathrm{diag}\{\mathbf{0}, \mathbf{G}_{p+1}, \cdots, \mathbf{G}_{p+q}\} \quad (6)$$

where $\mathbf{G}_i = \mathbf{G}_i^* \in \mathcal{C}^{k_i\times k_i}$ and $\mathcal{G}_{\mathbf{k}} \in \mathcal{C}^{n\times n}$, and

$$\begin{aligned}\alpha_{\mathbf{k}}(\mathbf{M}) \triangleq \inf_{\substack{\mathbf{D}\in\mathcal{D}_{\mathbf{k}}\\ \mathbf{G}\in\mathcal{G}_{\mathbf{k}}\\ 0<\alpha\in\mathcal{R}}} \{\alpha : \mathbf{M}^*\mathbf{D}\mathbf{M} \\ +\sqrt{-1}(\mathbf{G}\mathbf{M} - \mathbf{M}^*\mathbf{G}) - \alpha^2\mathbf{D} < 0\}. \quad (7)\end{aligned}$$

Then

$$\mu_{\mathbf{k}}(\mathbf{M}) \leq \alpha_{\mathbf{k}}(\mathbf{M}). \quad (8)$$

When $\mathbf{M}$ is a real matrix and the real scalars are not repeated, $\alpha_{\mathbf{k}}(\mathbf{M})$ can be computed easily. Define

$$\mathcal{D}_{\mathcal{R}\mathbf{k}} \triangleq \left\{\mathbf{D} \in \mathcal{D}_{\mathbf{k}} : \mathbf{D} \in \mathcal{R}^{n\times n}\right\}. \quad (9)$$

The following lemma is due to Young (Theorem 5.12 in [13]).

Lemma 4: Give a real matrix $\mathbf{M} \in \mathcal{R}^{n\times n}$ and a block structure $\mathbf{k}$ with $k_i = 1$ for $i = p+1, \cdots, p+q$, i.e. none of the real scalars are repeated. Then

$$\alpha_{\mathbf{k}}(\mathbf{M}) = \inf_{\substack{\mathbf{D}\in\mathcal{D}_{\mathcal{R}\mathbf{k}}\\ 0<\alpha\in\mathcal{R}}} \{\alpha : \mathbf{M}^{\mathrm{T}}\mathbf{D}\mathbf{M} - \alpha^2\mathbf{D} < 0\}. \quad (10)$$

Consider a matrix $\mathbf{M} \in \mathcal{C}^{n\times n}$ partitioned as

$$\mathbf{M} = \begin{bmatrix}\mathbf{M}_{11} & \mathbf{M}_{12}\\ \mathbf{M}_{21} & \mathbf{M}_{22}\end{bmatrix} \quad (11)$$

with $\mathbf{M}_{11} \in \mathcal{C}^{n_1\times n_1}, \mathbf{M}_{22} \in \mathcal{C}^{n_2\times n_2}$ and $n_1 + n_2 = n$. Suppose that we have block structure $\mathbf{k}_1$ and the corresponding perturbation set $\mathcal{K}_1$ compatible with $\mathbf{M}_{11}$, and block structure $\mathbf{k}_2$ and the corresponding perturbation set $\mathcal{K}_2$ compatible with $\mathbf{M}_{22}$. Then the block structure $\mathbf{k}_f \triangleq [\,\mathbf{k}_1^{\mathrm{T}} \quad \mathbf{k}_2^{\mathrm{T}}\,]^{\mathrm{T}}$ and the corresponding perturbation set

$$\mathcal{K}_f \triangleq \left\{\mathbf{\Upsilon} = \begin{bmatrix}\mathbf{\Upsilon}_1 & \\ & \mathbf{\Upsilon}_2\end{bmatrix} : \mathbf{\Upsilon}_1 \in \mathcal{K}_1, \mathbf{\Upsilon}_2 \in \mathcal{K}_2\right\} \quad (12)$$

is compatible with $\mathbf{M}$. Now given any $\mathbf{\Upsilon}_1 \in \mathcal{K}_1$,

$$F_u(\mathbf{M}, \mathbf{\Upsilon}_1) \triangleq \mathbf{M}_{22} + \mathbf{M}_{21}(\mathbf{I} - \mathbf{\Upsilon}_1\mathbf{M}_{11})^{-1}\mathbf{\Upsilon}_1\mathbf{M}_{12} \quad (13)$$

is called a linear fractional transformation (LFT) [14]. The main loop theorem for LFTs (Theorem 2.2 in [13]) is stated in the following lemma.

Lemma 5: Let $\mathbf{M} \in \mathcal{C}^{n\times n}$ and $0 < \alpha \in \mathcal{R}$. Then $\mu_{\mathbf{k}_f}(\mathbf{M}) < \alpha$ if and only if $\mu_{\mathbf{k}_1}(\mathbf{M}_{11}) < \alpha$, and for all $\mathbf{\Upsilon}_1 \in \mathcal{K}_1, \overline{\sigma}(\mathbf{\Upsilon}_1) \leq \frac{1}{\alpha}$ we have $\mu_{\mathbf{k}_2}(F_u(\mathbf{M}, \mathbf{\Upsilon}_1)) < \alpha$.

III. The FWL stability measure υ

Consider the discrete-time closed-loop control system consisting of a linear time-invariant plant $P(z)$ and a digital controller $C(z)$. The plant model $P(z)$ is assumed to be strictly proper with a state-space description

$$\begin{cases}\mathbf{x}_P(k+1) = \mathbf{A}_P\mathbf{x}_P(k) + \mathbf{B}_P\mathbf{u}(k)\\ \mathbf{y}(k) = \mathbf{C}_P\mathbf{x}_P(k)\end{cases} \quad (14)$$

where $\mathbf{A}_P \in \mathcal{R}^{r\times r}$, $\mathbf{B}_P \in \mathcal{R}^{r\times s}$ and $\mathbf{C}_P \in \mathcal{R}^{t\times r}$. The digital controller $C(z)$ is described by

$$\begin{cases}\mathbf{x}_C(k+1) = \mathbf{A}_C\mathbf{x}_C(k) + \mathbf{B}_C\mathbf{y}(k)\\ \mathbf{u}(k) = \mathbf{C}_C\mathbf{x}_C(k) + \mathbf{D}_C\mathbf{y}(k)\end{cases} \quad (15)$$

with $\mathbf{A}_C \in \mathcal{R}^{m\times m}$, $\mathbf{B}_C \in \mathcal{R}^{m\times t}$, $\mathbf{C}_C \in \mathcal{R}^{s\times m}$ and $\mathbf{D}_C \in \mathcal{R}^{s\times t}$. Denote the *realization* of $C(z)$ as

$$\mathbf{X} \triangleq \begin{bmatrix}\mathbf{D}_C & \mathbf{C}_C\\ \mathbf{B}_C & \mathbf{A}_C\end{bmatrix}. \quad (16)$$

Suppose that an initial realization of $C(z)$

$$\mathbf{X}_0 \triangleq \begin{bmatrix}\mathbf{D}_C^0 & \mathbf{C}_C^0\\ \mathbf{B}_C^0 & \mathbf{A}_C^0\end{bmatrix} \quad (17)$$

is given by some controller synthesis method. All the realizations of $C(z)$ form a set

$$\mathcal{X} \triangleq \left\{\mathbf{X} : \mathbf{X} = \mathbf{X}(\mathbf{T}) = \begin{bmatrix}\mathbf{I}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{T}\end{bmatrix}\mathbf{X}_0\begin{bmatrix}\mathbf{I}_t & \mathbf{0}\\ \mathbf{0} & \mathbf{T}^{-1}\end{bmatrix}\right\} \quad (18)$$

where the transformation $\mathbf{T} \in \mathcal{R}^{m\times m}$ is an arbitrary non-singular matrix. The stability of the closed-loop control system depends on the spectral radius of the closed-loop transition matrix

$$\begin{aligned}\overline{\mathbf{A}}(\mathbf{X}) &= \begin{bmatrix}\mathbf{A}_P + \mathbf{B}_P\mathbf{D}_C\mathbf{C}_P & \mathbf{B}_P\mathbf{C}_C\\ \mathbf{B}_C\mathbf{C}_P & \mathbf{A}_C\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{A}_P & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_P & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}\mathbf{X}\begin{bmatrix}\mathbf{C}_P & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{bmatrix}\\ &\triangleq \mathbf{M}_0 + \mathbf{M}_1\mathbf{X}\mathbf{M}_2. \quad (19)\end{aligned}$$

A discrete-time filter system can be viewed as a trivial case of the closed-loop system (14) and (15) with $P(z) = \mathbf{0}$, $r = 0$ and $C(z)$ representing the filter. Accordingly, the stability of the filter system depends on $\overline{\mathbf{A}}(\mathbf{X})$ with $\mathbf{M}_0 = \mathbf{0}$, $\mathbf{M}_1 = \mathbf{I}$, $\mathbf{M}_2 = \mathbf{I}$ and $\mathbf{X} = \mathbf{A}_C$, i.e. $\overline{\mathbf{A}}(\mathbf{X}) = \mathbf{A}_C$ as well as $\mathcal{X} = \{\mathbf{X} : \mathbf{X} = \mathbf{T}\mathbf{A}_C^0\mathbf{T}^{-1}\}$.

All the different realizations $\mathbf{X} \in \mathcal{X}$ have exactly the same set of poles if they are implemented with infinite precision. Since the system has been designed to be stable, $\rho(\overline{\mathbf{A}}(\mathbf{X})) < 1$. When $\mathbf{X}$ is implemented in an FWL fixed-point format, it is perturbed to $\mathbf{X}+\mathbf{\Delta}$. Each element of $\mathbf{\Delta}$ is bounded by $\pm\varepsilon$, that is, $\|\mathbf{\Delta}\|_{\rm m} \le \varepsilon$, where ε is the maximum representation error of the digital processor. With the perturbation $\mathbf{\Delta}$, $\overline{\mathbf{A}}(\mathbf{X})$ is moved to

$$\overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta}) = \overline{\mathbf{A}}(\mathbf{X}) + \mathbf{M}_1\mathbf{\Delta}\mathbf{M}_2. \tag{20}$$

If $\rho(\overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta})) \ge 1$, the system, designed to be stable, becomes unstable with the FWL implemented $\mathbf{X}$. It is therefore critical to know how robust the closed-loop stability to the FWL error $\mathbf{\Delta}$ for a realization $\mathbf{X} \in \mathcal{X}$. This means that we would like to know the largest open "hypercube" in the perturbation space within which the system remains stable. The size of this perturbation hypercube quantifies the FWL stability characteristics of $\mathbf{X}$ and is defined by the following FWL stability measure [10]

$$\upsilon(\mathbf{X}) \triangleq \inf_{\mathbf{\Delta}\in\mathcal{R}^{(s+m)\times(t+m)}} \{\|\mathbf{\Delta}\|_{\rm m} : \overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta}) \text{ is unstable}\}. \tag{21}$$

From the definition of $\upsilon(\mathbf{X})$, it is easy to see:

Theorem 1: $\overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta})$ is stable if $\|\mathbf{\Delta}\|_{\rm m} < \upsilon(\mathbf{X})$.

Theorem 1 implies that the larger $\upsilon(\mathbf{X})$ is, the larger FWL errors the realization $\mathbf{X}$ can tolerate. Moreover, as the FWL stability measure $\upsilon(\mathbf{X})$ is a function of $\mathbf{X}$, we can search for an "optimal" realization that maximizes $\upsilon(\mathbf{X})$

$$\mathbf{X}_{\rm opt} = \arg\max_{\mathbf{X}\in\mathcal{X}} \upsilon(\mathbf{X}). \tag{22}$$

The difficulty with this approach is that computing explicitly the value of $\upsilon(\mathbf{X})$ is still an unsolved open problem. In the next section, an SSV-based FWL stability measure is derived which not only can quantify the FWL effects on stability but can also be computed and optimized easily.

IV. An SSV-based FWL stability measure

Denote $N \triangleq (s+m)(t+m)$, and revisit (20) by defining

$$\begin{bmatrix} \mathbf{c}_1^{\rm T} & \cdots & \mathbf{c}_t^{\rm T} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{e}_1^{\rm T} & \cdots & \mathbf{e}_m^{\rm T} \end{bmatrix}^{\rm T} \triangleq \begin{bmatrix} \mathbf{C}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \mathbf{M}_2, \tag{23}$$

$$\begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1,t+m} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2,t+m} \\ \vdots & \vdots & \cdots & \vdots \\ \delta_{s+m,1} & \delta_{s+m,2} & \cdots & \delta_{s+m,t+m} \end{bmatrix} \triangleq \mathbf{\Delta}. \tag{24}$$

It is easy to check that

$$\overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta}) = \overline{\mathbf{A}}(\mathbf{X}) + \mathbf{B}_u\mathbf{\Lambda}\mathbf{C}_u \tag{25}$$

where

$$\mathbf{B}_u \triangleq [\overbrace{\mathbf{M}_1 \ \cdots \ \mathbf{M}_1}^{t+m}] \in \mathcal{R}^{(r+m)\times N}, \tag{26}$$

$$\mathbf{C}_u \triangleq \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix} \in \mathcal{R}^{N\times(r+m)}, \tag{27}$$

$$\mathbf{C}_1 \triangleq [\overbrace{\mathbf{c}_1^{\rm T} \ \cdots \ \mathbf{c}_1^{\rm T}}^{s+m} \ \cdots \ \overbrace{\mathbf{c}_t^{\rm T} \ \cdots \ \mathbf{c}_t^{\rm T}}^{s+m}]^{\rm T} \in \mathcal{R}^{(ts+tm)\times r}, \tag{28}$$

$$\mathbf{C}_2 \triangleq [\overbrace{\mathbf{e}_1^{\rm T} \ \cdots \ \mathbf{e}_1^{\rm T}}^{s+m} \ \cdots \ \overbrace{\mathbf{e}_m^{\rm T} \ \cdots \ \mathbf{e}_m^{\rm T}}^{s+m}]^{\rm T} \in \mathcal{R}^{(ms+m^2)\times m}, \tag{29}$$

$$\begin{aligned} \mathbf{\Lambda} \ \triangleq \ & \text{diag}\{\delta_{11}, \cdots, \delta_{s+m,1}, \delta_{12}, \cdots, \delta_{s+m,2}, \cdots, \\ & \delta_{1,t+m}, \cdots, \delta_{s+m,t+m}\} \in \mathcal{R}^{N\times N} \end{aligned} \tag{30}$$

with $\overline{\sigma}(\mathbf{\Lambda}) = \|\mathbf{\Delta}\|_{\rm m}$. For $0 < \beta \in \mathcal{R}$, denote

$$\mathbf{H}(\mathbf{X},\beta) \triangleq \begin{bmatrix} \overline{\mathbf{A}}(\mathbf{X}) & \mathbf{B}_u \\ \beta\mathbf{C}_u & \mathbf{0} \end{bmatrix} \in \mathcal{R}^{(r+m+N)\times(r+m+N)} \tag{31}$$

Choose $p_1 = 1$, $q_1 = 0$, block structure $\mathbf{k}_1(p_1,q_1) = r+m$, $p_2 = 0$, $q_2 = N$ and block structure $\mathbf{k}_2(p_2,q_2) = [\overbrace{1 \ \cdots \ 1}^{N}]^{\rm T}$. Clearly,

$$\mathcal{K}_1 = \{w\mathbf{I}_{r+m} : w \in \mathcal{C}\}, \tag{32}$$

$$\mathcal{K}_2 = \{\mathbf{\Lambda} \in \mathcal{R}^{N\times N} : \mathbf{\Lambda} \text{ is diagonal}\}, \tag{33}$$

$$\mathbf{k}_f = [r+m \ \overbrace{1 \ \cdots \ 1}^{N}]^{\rm T}. \tag{34}$$

Thus the perturbation set $\mathcal{K}_f$ given in (12) is compatible with $\mathbf{H}(\mathbf{X},\beta)$ and hence there exists $\mu_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta))$.

Theorem 2: $\upsilon(\mathbf{X}) > \beta > 0$ if and only if $\mu_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta)) < 1$.

Space limitation precludes the proof of Theorem 2.

Based on Theorem 2, it is easy to understand the relationship between $\upsilon(\mathbf{X})$ and $\mu_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta))$ as

Theorem 3: $\upsilon(\mathbf{X}) = \sup\{\beta \in \mathcal{R} : \beta > 0, \mu_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta)) < 1\}$.

Although we have successfully expressed $\upsilon(\mathbf{X})$ in the form of SSV, the difficulty in computing $\mu_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta))$ means that we have to explore a tractable lower bound of $\upsilon(\mathbf{X})$ with $\alpha_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta))$. Define

$$\mathcal{B} \triangleq \{\beta \in \mathcal{R} : \beta > 0, \alpha_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X},\beta)) < 1\}. \tag{35}$$

Some properties of $\mathcal{B}$ are now discussed. Since $\mathbf{H}(\mathbf{X},\beta)$ satisfies the conditions of Lemma 4, we have the following Theorem 4.

Theorem 4: $0 < \beta \in \mathcal{B}$ if and only if $\exists \mathbf{D} \in \mathcal{D}_{\mathcal{R}\mathbf{k}_f}$ such that

$$\mathbf{H}^{\rm T}(\mathbf{X},\beta)\mathbf{D}\mathbf{H}(\mathbf{X},\beta) - \mathbf{D} < 0. \tag{36}$$

Due to space limitation, we give the following two theorems without providing proofs.

Theorem 5: $\mathcal{B}$ is not empty.

Theorem 6: Suppose $\beta_1 > \beta_2 > 0$ and $\beta_1 \in \mathcal{B}$, then $\beta_2 \in \mathcal{B}$.

Now define

$$\upsilon_\mu(\mathbf{X}) \triangleq \sup_{\beta\in\mathcal{B}} \beta. \tag{37}$$

The following results based on (8) and Theorem 1 show that $\upsilon_\mu(\mathbf{X})$ can be viewed as an FWL stability measure which is a lower bound of $\upsilon(\mathbf{X})$.

Theorem 7: $\upsilon(\mathbf{X}) \geq \upsilon_\mu(\mathbf{X})$.

Theorem 8: $\overline{\mathbf{A}}(\mathbf{X}+\mathbf{\Delta})$ is stable if $\|\mathbf{\Delta}\|_{\rm m} < \upsilon_\mu(\mathbf{X})$.

Through the discussion for $\mathcal{B}$, we know that

$$\mathcal{B} \cup \{\upsilon_\mu\} = (0, \upsilon_\mu] \tag{38}$$

is non-empty and bounded (The fact that $\upsilon(\mathbf{X})$ is finitely large implies $\upsilon_\mu(\mathbf{X}) \neq \infty$). Therefore, given a realization $\mathbf{X}$, one can compute $\upsilon_\mu(\mathbf{X})$ conveniently based on the following bisection searching.

Step 1 Determine a precision $\tau > 0$. Initially set a small enough β_1 such that $\beta_1 \in \mathcal{B}$ and a large enough β_2 such that $\beta_2 \notin \mathcal{B}$.

Step 2 Set $\beta_3 = (\beta_1+\beta_2)/2$, and solve the LMI

$$\begin{bmatrix} \overline{\mathbf{A}}(\mathbf{X}) & \mathbf{B}_u \\ \beta_3\mathbf{C}_u & \mathbf{0} \end{bmatrix}^{\rm T} \mathbf{D} \begin{bmatrix} \overline{\mathbf{A}}(\mathbf{X}) & \mathbf{B}_u \\ \beta_3\mathbf{C}_u & \mathbf{0} \end{bmatrix} - \mathbf{D} < 0$$

$$0 < \mathbf{D} \in \mathcal{D}_{\mathcal{R}\mathbf{k}_f}$$

with the LMI toolbox of MATLAB.

Step 3 If the above LMI has a solution, let $\beta_1 = \beta_3$; if the LMI has no solution, let $\beta_2 = \beta_3$.

Step 4 If $\beta_2 - \beta_1 < \tau$, Let $\upsilon_\mu(\mathbf{X}) = \beta_1$ and terminate the algorithm; if $\beta_2 - \beta_1 \geq \tau$, go to *Step 2*.

V. Optimal FWL realizations

The SSV-based stability measure $\upsilon_\mu(\mathbf{X})$ is a function of the realization $\mathbf{X}$. It is of practical importance to find an "optimal" realization that maximizes $\upsilon_\mu(\mathbf{X})$ over $\mathcal{X}$. The filter/controller implemented with this realization can tolerate a maximum FWL error. Since $\mathbf{X} \in \mathcal{X}$ depends on the non-singular transformation matrix $\mathbf{T}$, the optimal FWL realization problem is formally defined as

$$\gamma \triangleq \sup_{\substack{\mathbf{T}\in\mathcal{R}^{m\times m}\\ \det\mathbf{T}\neq 0}} \upsilon_\mu(\mathbf{X}(\mathbf{T})). \tag{39}$$

Combining (35), (37) and (39), we have

$$\gamma = \sup_{\substack{\mathbf{T}\in\mathcal{R}^{m\times m}\\ \det\mathbf{T}\neq 0\\ 0<\beta\in\mathcal{R}}} \{\beta : \alpha_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X}(\mathbf{T}),\beta)) < 1\}. \tag{40}$$

We now show how the optimal realization problem (40) can be solved using the LMI technique. Let $0 < \mathbf{P}_1 \in \mathcal{R}^{(r+m)\times(r+m)}$, $0 < \mathbf{P}_2 \in \mathcal{R}^{r\times r}$, $0 < \mathbf{P}_3 \in \mathcal{R}^{m\times m}$, $0 < v_i \in \mathcal{R}$, $i \in \{1,\cdots,N\}$ and $\mathbf{T} \in \mathcal{R}^{m\times m}$. First define

$$\mathbf{G}_{1,1} = \mathbf{P}_1 - \overline{\mathbf{A}}^{\rm T}(\mathbf{X}_0)\mathbf{P}_1\overline{\mathbf{A}}(\mathbf{X}_0), \tag{41}$$

$$\mathbf{G}_{2,1} = -\mathbf{B}_u^{\rm T}\mathbf{P}_1\overline{\mathbf{A}}(\mathbf{X}_0), \tag{42}$$

$$\mathbf{G}_{2,2} = \begin{bmatrix} \begin{bmatrix} 2\mathbf{I}_s & \\ & \mathbf{T}^{\rm T}+\mathbf{T} \end{bmatrix} & & \\ & \ddots & \\ & & \begin{bmatrix} 2\mathbf{I}_s & \\ & \mathbf{T}^{\rm T}+\mathbf{T} \end{bmatrix} \end{bmatrix}, \tag{43}$$

$$\mathbf{Q}_j \triangleq \mathrm{diag}\{v_j, v_{(s+m)+j}, \cdots, v_{(t-1)(s+m)+j}\} \tag{44}$$

with $j \in \{1,\cdots,s+m\}$ and

$$\mathbf{W}_j \triangleq \mathrm{diag}\{v_{t(s+m)+j}, v_{(t+1)(s+m)+j}, \cdots, v_{(t+m-1)(s+m)+j}\} \tag{45}$$

with $j \in \{1,\cdots,s+m\}$. Next introduce the following LMIs

$$\left[\begin{array}{c|c} \mathbf{G}_{1,1} & \sharp \\ \hline \mathbf{G}_{2,1} & \mathbf{G}_{2,2} \end{array}\right] >$$

$$\left[\begin{array}{cc|c} \beta^2\mathbf{P}_2 & & \\ & \beta^2\mathbf{P}_3 & \\ \hline & & \mathbf{B}_u^{\rm T}\mathbf{P}_1\mathbf{B}_u + \begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_N \end{bmatrix} \end{array}\right], \tag{46}$$

$$\begin{bmatrix} \mathbf{P}_2 & \mathbf{C}_P^{\rm T} & \cdots & \mathbf{C}_P^{\rm T} \\ \mathbf{C}_P & \mathbf{Q}_1 & & \\ \vdots & & \ddots & \\ \mathbf{C}_P & & & \mathbf{Q}_{s+m} \end{bmatrix} > 0, \tag{47}$$

$$\begin{bmatrix} \mathbf{P}_3 & \mathbf{T}^{\rm T} & \cdots & \mathbf{T}^{\rm T} \\ \mathbf{T} & \mathbf{W}_1 & & \\ \vdots & & \ddots & \\ \mathbf{T} & & & \mathbf{W}_{s+m} \end{bmatrix} > 0. \tag{48}$$

Theorem 9: Suppose that for a positive $\beta \in \mathcal{R}$ the LMI (46)–(48) has a solution, that is, there exist $0 < \mathbf{P}_1 \in \mathcal{R}^{(r+m)\times(r+m)}$, $0 < \mathbf{P}_2 \in \mathcal{R}^{r\times r}$, $0 < \mathbf{P}_3 \in \mathcal{R}^{m\times m}$, $0 < v_i \in \mathcal{R}$, $i \in \{1,\cdots,N\}$ and $\mathbf{T} \in \mathcal{R}^{m\times m}$ such that the LMI (46)–(48) holds. Then $\alpha_{\mathbf{k}_f}(\mathbf{H}(\mathbf{X}(\mathbf{T}),\beta)) < 1$.

Again the proof of Theorem 9 is omitted owing to space limitation.

Let us now define

$$\mathcal{B}_{\mathbf{T}} \triangleq \{\beta \in \mathcal{R} : \beta > 0, \text{LMI (46)–(48) has a solution}\}. \tag{49}$$

It is easy to prove that $\mathcal{B}_{\mathbf{T}}$ has the similar properties to those of $\mathcal{B}$ as described in Theorems 5 and 6. Therefore, we can solve the optimal FWL realization problem in the following procedure.

Step 1 Determine a precision $\tau > 0$. Initially set a small enough β_1 such that $\beta_1 \in \mathcal{B}_{\mathbf{T}}$ and a large enough β_2 such that $\beta_2 \notin \mathcal{B}_{\mathbf{T}}$.

Step 2 Set $\beta_3 = (\beta_1+\beta_2)/2$, and solve the LMI (46)–(48).

Step 3 If the above LMI has a solution, set $\beta_1 = \beta_3$ and $\mathbf{T}_{\rm opt} = \mathbf{T}$; if the LMI has no solution, set $\beta_2 = \beta_3$.

Step 4 If $\beta_2 - \beta_1 < \tau$, go to *Step 5*; if $\beta_2 - \beta_1 \geq \tau$, go to *Step 2*.

Step 5 The optimal realization is

$$\mathbf{X}_{\rm opt} = \begin{bmatrix} \mathbf{I}_s & \\ & \mathbf{T}_{\rm opt} \end{bmatrix} \mathbf{X}_0 \begin{bmatrix} \mathbf{I}_t & \\ & \mathbf{T}_{\rm opt}^{-1} \end{bmatrix}. \qquad (50)$$

Use the search algorithm given in Section IV to compute $\tilde{\gamma} = \upsilon_\mu(\mathbf{X}_{\rm opt})$.

Comment: From the proof of Theorem 9 (which was not included), it can be seen that the algorithm presented here is slightly conservative and in general $\tilde{\gamma}$ is less than the true maximum γ. However, we can obtain a satisfactory realization $\mathbf{X}_{\rm opt}$ whose $\upsilon_\mu(\mathbf{X}_{\rm opt})$ at least is larger than $\sup_{\beta \in \mathcal{B}_{\mathbf{T}}} \beta$.

VI. A NUMERICAL EXAMPLE

The plant was defined by

$$\mathbf{A}_P = \begin{bmatrix} 9.9513e-1 & -9.7260 & 4.8724e-3 \\ 9.9614e-4 & 9.8843e-1 & -9.9614e-4 \\ 6.6995e-3 & 1.3373e1 & 9.9330e-1 \end{bmatrix},$$

$$\mathbf{B}_P = \begin{bmatrix} 2.4863e-1 \\ 1.2427e-4 \\ 5.5656e-4 \end{bmatrix},\ \mathbf{C}_P = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^{\rm T}$$

and an initial realization of the controller was given by

$$\mathbf{X}_0 = \left[\begin{array}{c|cc} 1.3512 & 1.4260e-2 & 1.1956 \\ \hline -1 & 1 & 0 \\ -1 & 0 & 3.3330e-1 \end{array}\right].$$

The value of the SSV-based stability measure for this initial controller realization was computed by the algorithm given in Section IV as $\upsilon_\mu(\mathbf{X}_0) = 4.3241e-3$. Using the method presented in Section V, we obtained the optimal FWL transformation matrix

$$\mathbf{T}_{\rm opt} = \begin{bmatrix} 1.0993e-1 & -1.0858e-1 \\ 2.4484e-2 & 1.0785 \end{bmatrix}$$

and computed the optimal FWL controller realization as

$$\mathbf{X}_{\rm opt} = \left[\begin{array}{c|cc} 1.3512 & -1.1460e-1 & 1.0970 \\ \hline -1.3490e-3 & 9.8538e-1 & 6.5647e-2 \\ -1.1030 & 1.4523e-1 & 3.4792e-1 \end{array}\right].$$

The value of the SSV-based stability measure for this optimal realization was $\upsilon_\mu(\mathbf{X}_{\rm opt}) = 1.3128e-2$, which is three times the value for the initial realization.

VII. CONCLUSIONS

Based on the structured singular value theory, a computationally tractable stability measure has been derived for the digital controller/filter with FWL implementation considerations. The optimal FWL realization problem for the controller/filter has been defined based on this SSV-based stability measure, and an efficient optimization strategy has been presented to solve this optimization problem using the LMI technique. A numerical example has been included to illustrate the proposed FWL design procedure.

ACKNOWLEDGEMENTS

This work is partially supported by Tan Chin Tuan exchange fellowship, NSFC #60374002, #60421002, 973 program of China #2002CB312200, NCET-04-0547 and UK Royal Academy of Engineering.

REFERENCES

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

[2] R.S.H. Istepanian and J.F. Whidborne, eds., *Digital Controller Implementation and Fragility: A Modern Perspective*. London: Springer Verlag, 2001.

[3] L.H. Keel and S.P. Bhattacharyya, "Stability margins and digital implementation of controllers," in: R.S.H. Istepanian and J.F. Whidborne, eds., *Digital Controller Implementation and Fragility: A Modern Perspective*, London: Springer Verlag, 2001 pp.13–24.

[4] G. Li, "On the structure of digital controllers with finite word length consideration," *IEEE Trans. Automatic Control*, Vol.43, No.5, pp.689–693, 1998.

[5] J.F. Whidborne, J. Wu and R.S.H. Istepanian, "Finite word length stability issues in an l_1 framework," *Int. J. Control*, Vol.73, No.2, pp.166–176, 2000.

[6] P. Mantey, "Eigenvalue sensitivity and state-variable selection," *IEEE Trans. Automatic Control*, Vol.13, No.3, pp.263–269, 1968.

[7] J. Wu, S. Chen, G. Li, R.S.H. Istepanian and J. Chu, "An improved closed-loop stability related measure for finite-precision digital controller realizations," *IEEE Trans. Automatic Control*, Vol.46, No.7, pp.1162–1166, 2001.

[8] I.J. Fialho and T.T. Georgiou, "Computational algorithms for sparse optimal digital controller realizations," in: R.S.H. Istepanian and J.F. Whidborne, eds., *Digital Controller Implementation and Fragility: A Modern Perspective*, London: Springer Verlag, 2001, pp.105–121.

[9] W. Yu and H. Ko, "Improved eigenvalue sensitivity for finite-precision digital controller realisations via orthogonal Hermitian transform," *IEE Proc. Control Theory and Applications*, Vol.150, No.4, pp.365–375, 2003.

[10] I.J. Fialho and T.T. Georgiou, "On stability and performance of sampled-data systems subject to wordlength constraint," *IEEE Trans. Automatic Control*, Vol.39, No.12, pp.2476–2481, 1994.

[11] M.A. Dahleh and I.J. Diaz-Bobillo, *Control of Uncertain Systems: A Linear Programming Approach*. NJ: Prentice-Hall, 1995.

[12] J. Doyle, "Analysis of feedback systems with structured uncertainty," *IEE Proc. Control Theory and Applications*, Vol.129, No.6, pp.242–250, 1982.

[13] P.M. Young, *Robustness with Parametric and Dynamic Uncertainty*. PhD thesis, California Institute of Technology, 1993.

[14] J. Doyle, A. Packard and K. Zhou, "Review of LFTs, LMIs and μ," in: *Proc. 30th Conf. Decision and Control* (Brighton, England), 1991, pp.1227–1232.