Sparse Realizations of Optimal Finite-Precision Teleoperation Controller Structures

R.H. Istepanian[†], *J. Wu*[‡] *and S. Chen*[¶]

† Department of Electrical and Computer Engineering
Ryerson Polytechnic University, Toronto, Ontario, Canada M5B 2K3

‡ Institute of Industrial Process Control
Zhejiang University, Hangzhou, 310027, P.R. China

¶ Department of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, U.K.

Abstract

We study the finite word length (FWL) implementation of digital controller structures with sparseness consideration. A FWL stability measure is derived, taking into account the number of trivial elements in a controller realization. The controller realization that maximizes a lower bound of this measure is first obtained, and a stepwise algorithm is then applied to make the realization sparse. A test case involving a dual wrist assembly shows that the proposed design procedure yields a computationally efficient controller realization with good FWL closed-loop stability performance.

1 Introduction

In real-time applications where computational efficiency is critical, a digital controller implemented in fixed-point arithmetic has certain advantages. However, a stable control system may achieve a lower than predicted performance or even become unstable when the control law is implemented with a fixed-point device due to the FWL effects. The FWL effects on the closed-loop stability depend on the controller realization structure. Earlier studies have addressed the problem of finding the "optimal" realization of finite-precision controller structures, which has a maximum tolerance to FWL errors but may not have a sparse structure [1, 2].

It is highly desirable that a controller realization has a sparse structure, containing many trivial elements of 0, 1 or -1. It is known that canonical controller realizations have sparse structures but may not have the required FWL stability robustness. An optimal controller realization that maximizes the stability measures of [1, 2] usually is a fully parameterized structure. A sparse structure is particularly important for real-time applications with high-order controllers, as it will achieve better computational efficiency. This poses a complex problem of finding sparse controller realizations with good FWL closed-loop stability characteristics.

We present an FWL stability measure, taking into account the sparseness consideration. The true optimal realization that maximizes this measure will possess an optimal trade-off between robustness to FWL errors and sparse structure. Unfortunately, it is not known how to obtain such an optimal realization. We extend an iterative algorithm [3, 4] to search for a suboptimal solution. Specifically, we first obtain the realization that maximizes a lower bound of the proposed stability measure. This can easily be done [1, 2] but the resulting realization is not sparse. A stepwise algorithm is then applied to make the realization sparse without reducing the FWL stability measure too much.

The design procedure is tested on a dual wrist assembly, which is a prototype telerobotic system used in microsurgery experiments [5]. This dual wrist assembly is a two-input two-output system with a plant order of 4, and the digital controller designed using H_∞ method has an order of 10 [5]. The total number of controller parameters is 144. As this controller is a high-order one and fast sampling is used, a sparseness realization with good FWL stability characteristics is crucial for computational efficiency in real-time operation.

2 A FWL stability measure with sparseness consideration

Consider the discrete-time control system with plant $P(z)$ and controller $C(z)$, depicted in Fig. 1. $P(z)$ is assumed to be strictly causal. Let $(A_z, B_z, C_z, 0)$ be a state-space description of $P(z)$ with $A_z \in R^{m\times m}$, $B_z \in$

$R^{m\times l}$ and $C_z \in R^{q\times m}$, and (A_c, B_c, C_c, D_c) be a state-space description of $C(z)$ with $A_c \in R^{n\times n}$, $B_c \in R^{n\times q}$, $C_c \in R^{l\times n}$ and $D_c \in R^{l\times q}$. Then the stability of the closed-loop control system depends on the poles of the closed-loop system matrix

$$\overline{A} = \left[\begin{array}{cc} A_z + B_z D_c C_z & B_z C_c \\ B_c C_z & A_c \end{array} \right] \tag{1}$$

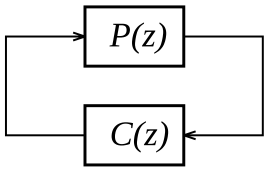

Figure 1: Discrete-time control system consisting of plant $P(z)$ and controller $C(z)$.

Any linear system with a given transfer function matrix has an infinite number of state-space descriptions. In fact, if $(A_c^0, B_c^0, C_c^0, D_c^0)$ is a state-space description of the digital controller $C(z)$, all the state-space descriptions of $C(z)$ form a set

$$\begin{aligned} \mathcal{S}_C \;\; &\triangleq \;\; \{(A_c, B_c, C_c, D_c) : A_c = T^{-1}A_c^0 T, \\ & \quad\; B_c = T^{-1}B_c^0, C_c = C_c^0 T, D_c = D_c^0\} \end{aligned} \tag{2}$$

where $T \in R^{n\times n}$ is any nonsingular matrix, called a similarity transformation. Any $(A_c, B_c, C_c, D_c) \in \mathcal{S}_C$ is a *realization* of $C(z)$. Denote $N \triangleq (l+n)(q+n)$ and

$$X \triangleq \left[\begin{array}{cc} D_c & C_c \\ B_c & A_c \end{array} \right] = \left[\begin{array}{ccc} p_1 & \cdots & p_{N-l-n+1} \\ \vdots & \cdots & \vdots \\ p_{l+n} & \cdots & p_N \end{array} \right] \tag{3}$$

We will also refer to X as a realization of $C(z)$. From (1), we know that $\overline{A}$ is a function of X

$$\begin{aligned} \overline{A}(X) &= \left[\begin{array}{cc} A_z & 0 \\ 0 & 0 \end{array} \right] + \left[\begin{array}{cc} B_z & 0 \\ 0 & I \end{array} \right] X \left[\begin{array}{cc} C_z & 0 \\ 0 & I \end{array} \right] \\ &\triangleq M_0 + M_1 X M_2 \end{aligned} \tag{4}$$

When the fixed-point format is used to implement the controller, X is perturbed into $X + \Delta X$, where

$$\Delta X \triangleq \left[\begin{array}{ccc} \Delta p_1 & \cdots & \Delta p_{N-l-n+1} \\ \vdots & \cdots & \vdots \\ \Delta p_{l+n} & \cdots & \Delta p_N \end{array} \right] \tag{5}$$

and each element of ΔX is bounded by $\varepsilon/2$ such that

$$\mu(\Delta X) \triangleq \max_{i\in\{1,\cdots,N\}} |\Delta p_i| \le \frac{\varepsilon}{2} \tag{6}$$

Obviously, $\mu(\Delta X)$ is a norm of the FWL error ΔX. For a fixed-point processor of B_s bits

$$\varepsilon = 2^{-(B_s - B_X)} \tag{7}$$

where 2^{B_X} is a normalization factor. With the perturbation ΔX, a closed-loop pole $\lambda_i(\overline{A}(X))$ of the originally stable system is moved to $\lambda_i(\overline{A}(X + \Delta X))$, which may be outside the open unit disk and hence causes the closed-loop to become unstable.

Note that the parameters 0, 1 and -1 are *trivial*, since they require no operations in the fixed-point implementation and do not cause any computation error at all. Thus $\Delta p_i = 0$ when $p_i = 0$, 1 or -1. Let us define

$$\delta(p) = \begin{cases} 0, & \text{if } p = 0, 1 \text{ or } -1 \\ 1, & \text{otherwise} \end{cases} \tag{8}$$

When ΔX is small, we notice that

$$\begin{aligned} \Delta\lambda_i &\triangleq \lambda_i(\overline{A}(X + \Delta X)) - \lambda_i(\overline{A}(X)) \\ &\approx \sum_{j=1}^{N} \frac{\partial \lambda_i}{\partial p_j} \Delta p_j \delta(p_j), \;\; \forall i \in \{1, \cdots, m+n\} \end{aligned} \tag{9}$$

It follows that

$$\begin{aligned} |\Delta\lambda_i| &\le \sqrt{N_s \sum_{j=1}^{N} \left| \frac{\partial \lambda_i}{\partial p_j} \right|^2 |\Delta p_j|^2 \delta(p_j)} \\ &\le \mu(\Delta X) \sqrt{N_s \sum_{j=1}^{N} \left| \frac{\partial \lambda_i}{\partial p_j} \right|^2 \delta(p_j)}, \;\; \forall i \end{aligned} \tag{10}$$

where N_s is the number of the nontrivial elements in X. Define

$$\mu_1(X) = \min_{i\in\{1,\cdots,m+n\}} \frac{1 - |\lambda_i(\overline{A}(X))|}{\sqrt{N_s \sum_{j=1}^{N} \delta(p_j) \left| \frac{\partial \lambda_i}{\partial p_j} \right|^2}} \tag{11}$$

If $\mu(\Delta X) < \mu_1(X)$, it follows from (10) and (11) that $|\Delta\lambda_i| < 1 - |\lambda_i(\overline{A}(X))|$. Therefore

$$|\lambda_i(\overline{A}(X + \Delta X))| \le |\Delta\lambda_i| + |\lambda_i(\overline{A}(X))| < 1 \tag{12}$$

which means that the closed-loop system remains stable under perturbation ΔX. In other words, the larger $\mu_1(X)$ is, the bigger FWL error that the closed-loop system can tolerate. Hence $\mu_1(X)$ is a measure describing the FWL stability characteristics of X.

The measure (11) takes into account the sparseness consideration. Furthermore, it is computationally tractable, as shown in the following theorem (see [2] for a proof).

Theorem 1 Let $\{\lambda_i\} = \{\lambda_i(\overline{A}(X))\}$ be the eigenvalues of $\overline{A}(X) = M_0 + M_1 X M_2$. Denote x_i and y_i the right and reciprocal left eigenvectors corresponding to λ_i, respectively. Then

$$\frac{\partial \lambda_i}{\partial X} = \left[\begin{array}{ccc} \frac{\partial \lambda_i}{\partial p_1} & \cdots & \frac{\partial \lambda_i}{\partial p_{N-l-n+1}} \\ \vdots & \cdots & \vdots \\ \frac{\partial \lambda_i}{\partial p_{l+n}} & \cdots & \frac{\partial \lambda_i}{\partial p_N} \end{array} \right] = M_1^{\mathcal{T}} y_i^* x_i^{\mathcal{T}} M_2^{\mathcal{T}} \quad (13)$$

where the superscript $*$ denotes the conjugate operation and $\mathcal{T}$ the transpose operation.

Let $B_s^{\min}$ be the smallest word length that, when used to implement X, can guarantee the closed-loop stability. An estimate of $B_s^{\min}$ is given by

$$\hat{B}_s^{\min} = \text{Int}[-\log_2(\mu_1(X))] - 1 + B_X \quad (14)$$

where $\text{Int}[x]$ rounds x to the nearest integer and $\text{Int}[x] \geq x$. The optimal sparse controller realization with a maximum tolerance to FWL perturbation in principle is the solution of the optimization problem

$$v \triangleq \max_{X \in \mathcal{S}_C} \mu_1(X) \quad (15)$$

However, we do not know how to solve the above problem because $\mu_1(X)$ includes $\delta(p_j)$ and is not a continuous function with respect to controller elements p_j.

3 Sparse realizations with good FWL stability characteristics

Consider a lower bound of $\mu_1(X)$

$$\mu_{1l}(X) = \min_{i \in \{1,\cdots,m+n\}} \frac{1 - \left|\lambda_i(\overline{A}(X))\right|}{\sqrt{N \sum_{j=1}^{N} \left|\frac{\partial \lambda_i}{\partial p_j}\right|^2}} \quad (16)$$

In fact, (16) is the stability measure given in [1], which does not take into account the number of trivial parameters in X but is a continuous function. Obviously, $\mu_{1l}(X) \leq \mu_1(X)$. The "optimal" realization that maximizes μ_{1l} is the solution of the following problem

$$\omega \triangleq \max_{X \in \mathcal{S}_C} \mu_{1l}(X) \quad (17)$$

and is relatively easy to obtain [1, 2] via the following optimization procedure.

Assume that an initial controller realization is given as

$$X_0 = \left[\begin{array}{cc} D_c^0 & C_c^0 \\ B_c^0 & A_c^0 \end{array} \right] \quad (18)$$

From (2) and (4), we have

$$X = X(T) = \left[\begin{array}{cc} I & 0 \\ 0 & T^{-1} \end{array} \right] X_0 \left[\begin{array}{cc} I & 0 \\ 0 & T \end{array} \right] \quad (19)$$

and

$$\overline{A}(X) = \left[\begin{array}{cc} I & 0 \\ 0 & T^{-1} \end{array} \right] \overline{A}(X_0) \left[\begin{array}{cc} I & 0 \\ 0 & T \end{array} \right] \quad (20)$$

Obviously, $\overline{A}(X)$ has the same eigenvalues as $\overline{A}(X_0)$, denoted as $\{\lambda_i^0\}$. Applying theorem 1 to (20) results in

$$\left.\frac{\partial \lambda_i}{\partial X}\right|_{X(T)} = \left[\begin{array}{cc} I & 0 \\ 0 & T^{\mathcal{T}} \end{array} \right] \left.\frac{\partial \lambda_i}{\partial X}\right|_{X_0} \left[\begin{array}{cc} I & 0 \\ 0 & T^{-\mathcal{T}} \end{array} \right] \quad (21)$$

For a complex-valued matrix $M \in C^{(l+n)\times(q+n)}$ with elements m_{ij}, define the Frobenius norm

$$\|M\|_F \triangleq \sqrt{\sum_{i=1}^{l+n} \sum_{j=1}^{q+n} m_{ij}^* m_{ij}} \quad (22)$$

Then the optimization problem (17) is equivalent to

$$\omega = \max_{\substack{T \in R^{n\times n} \\ \det(T) \neq 0}} f(T) \quad (23)$$

with the cost function

$$f(T) = \min_{i \in \{1,\cdots,m+n\}} \frac{1}{\sqrt{N} \left\| \left[\begin{array}{cc} I & 0 \\ 0 & T^{\mathcal{T}} \end{array} \right] \Phi_i \left[\begin{array}{cc} I & 0 \\ 0 & T^{-\mathcal{T}} \end{array} \right] \right\|_F} \quad (24)$$

where

$$\Phi_i \triangleq \frac{\left.\frac{\partial \lambda_i}{\partial X}\right|_{X=X_0}}{1 - |\lambda_i^0|} \quad (25)$$

The optimal similarity matrix T_{opt} can be obtained by solving for the unconstrained optimization problem

$$\omega = \max_{T \in R^{n\times n}} f(T) \quad (26)$$

with a measure of monitoring the singular values of T to make sure that $\det(T) \neq 0$. In this study, we use the simulated annealing [6] to solve for the unconstrained optimization problem (26). The corresponding controller realization is given by $X(T_{\text{opt}}) = X_{\text{opt}}$.

Notice that X_{opt} is not the optimal solution of the problem (15) and does not have a sparse structure. We can make X_{opt} sparse by changing one nontrivial element of X_{opt} into a trivial one at a step, under the constraint that the value of μ_{1l} does not reduce too much. This process will produce a suboptimal sparse realization X_{sop}, and the detailed stepwise algorithm is:

Step 1: Set τ to a very small positive real number (e.g. 10^{-5}). The transformation matrix T is initially set to T_{opt} so that $X(T) = X_{\text{opt}}$.

Step 2: Find out all the trivial elements $\{\eta_1, \cdots, \eta_r\}$ in $X(T)$ (a parameter is considered to be trivial if its distance from 0, 1 or -1 is less than 10^{-8}). Denote ξ the non-trivial element in $X(T)$ that is the nearest to 0, 1 or -1.

Step 3: Choose $S \in R^{n\times n}$ such that

i) $\mu_{1l}(X(T + \tau S))$ is close to $\mu_{1l}(X(T))$.

ii) $\{\eta_1, \cdots, \eta_r\}$ in $X(T)$ remain unchanged in $X(T + \tau S)$.

iii) ξ in $X(T)$ is changed to as near to 0, 1 or -1 as possible in $X(T + \tau S)$.

iv) $\|S\|_F = 1$.

If S does not exist, $T_{\text{sop}} = T$ and terminate the algorithm.

Step 4: $T = T + \tau S$. If ξ in $X(T)$ is non-trivial, go to step 3. If ξ becomes trivial, go to step 2.

The **step 3** is the key to guarantee that $X_{\text{sop}} = X(T_{\text{sop}})$ contains many trivial elements and has good performance as measured by μ_{1l}. We now discuss how to obtain S. First, denote $Vec(S)$ the vector containing the columns of the matrix S stacked in column order. With a very small τ, condition i) means

$$\left(Vec\left(\frac{d\mu_{1l}}{dT}\right)\right)^{\mathcal{T}} Vec(S) = 0 \tag{27}$$

Condition ii) means

$$\begin{cases} \left(Vec\left(\frac{d\eta_1}{dT}\right)\right)^{\mathcal{T}} Vec(S) = 0 \\ \qquad\vdots \\ \left(Vec\left(\frac{d\eta_r}{dT}\right)\right)^{\mathcal{T}} Vec(S) = 0, \end{cases} \tag{28}$$

Denote the matrix

$$E \triangleq \begin{bmatrix} \left(Vec\left(\frac{d\mu_{1l}}{dT}\right)\right)^{\mathcal{T}} \\ \left(Vec\left(\frac{d\eta_1}{dT}\right)\right)^{\mathcal{T}} \\ \vdots \\ \left(Vec\left(\frac{d\eta_r}{dT}\right)\right)^{\mathcal{T}} \end{bmatrix} \in R^{(r+1)\times n^2} \tag{29}$$

$Vec(S)$ must belong to the null space $\mathcal{N}(E)$ of E. If $\mathcal{N}(E)$ is empty, $Vec(S)$ does not exist and the algorithm is terminated. If $\mathcal{N}(E)$ is not empty, it must have basis $\{e_1, \cdots, e_t\}$, assuming that the dimension of $\mathcal{N}(E)$ is t. Condition iii) requires moving ξ closer to its desired value (0, 1 or -1) as fast as possible, and we should choose $Vec(S)$ as the orthogonal projection of $Vec\left(\frac{d\xi}{dT}\right)$ onto $\mathcal{N}(E)$. Noting condition iv), we can compute $Vec(S)$ as follows

$$a_i = e_i^{\mathcal{T}} Vec\left(\frac{d\xi}{dT}\right) \in R,\ \forall i \in \{1, \cdots, t\} \tag{30}$$

$$w = \sum_{i=1}^{t} a_i e_i \in R^{n^2} \tag{31}$$

$$Vec(S) = \pm \frac{w}{\sqrt{w^{\mathcal{T}} w}} \in R^{n^2} \tag{32}$$

The sign in (32) is chosen in the following way. If ξ is larger than its nearest desired value, the minus sign is taken; otherwise, the plus sign is used.

4 The dual wrist assembly case

A MATLAB program implementing the above algorithm was applied to the dual wrist assembly. The initial controller realization X_{ini}, chosen to be the H_∞ controller given in [5], had a low-bound stability measure of $\mu_{1l}(X_{\text{ini}}) = 1.1734 \times 10^{-4}$. The simulated annealing algorithm obtained X_{opt} with $\mu_{1l}(X_{\text{opt}}) = 1.5844 \times 10^{-3}$, and the stepwise algorithm made X_{opt} sparse to produce X_{sop} with $\mu_{1l}(X_{\text{sop}}) = 4.3325 \times 10^{-4}$. Table 1 summarizes the performance of these three different controller realizations. Notice that, although the algorithm operates based on μ_{1l}, the FWL stability characteristics are judged using μ_1 in Table 1.

Realization	μ_1	$\hat{B}_s^{\min}$	N_s
X_{ini}	1.1734×10^{-4}	29	144
X_{opt}	1.5844×10^{-3}	25	144
X_{sop}	1.1171×10^{-3}	25	63

Table 1: Comparison of different realizations.

It can be seen that, for this teleoperation system, both X_{opt} and X_{sop} can guarantee the closed-loop stability when implemented using a fixed-point processor of 25 bits while X_{ini} requires 29 bits. Although the value of stability measure is smaller for X_{sop} compared with X_{opt}, it has 81 trivial elements, out of the total of 144 parameters. Thus this sparse controller realization yields a computationally effective structure while maintaining good FWL closed-loop stability robustness.

Fig. 2 compares the closed-loop force tracking errors from the active operator hand force of the dual wrist assembly when the different controller realizations were implemented with a 30-bit processor. It can be seen from Fig. 2 that there is a clear difference between the performance of X_{ini} with FWL implementation and that of the ideal controller implemented with infinite bits. The 30-bit implemented X_{opt} and X_{sop}, however, produced the responses very closed to that of the ideal controller.

5 Conclusions

We have investigated the problem of digital controller implementations with FWL and sparseness considerations. A FWL closed-loop stability measure has been derived, which takes into account the number of trivial parameters in a controller realization. A practical stepwise procedure has been presented to obtain sparse controller realizations with satisfactory FWL closed-loop stability characteristics. A case study involving a teleoperation system with a high-order controller demonstrates that the proposed design procedure yields computationally efficient controller structures suitable for FWL implementation in real-time applications.

Acknowledgements

The authors thank Dr J. Yan, Department of Electrical and Computer Engineering, University of California at Berkeley, USA, for providing the model and controller of the dual wrist assembly.

References

[1] G. Li, "On the structure of digital controllers with finite word length consideration," *IEEE Trans. Automatic Control*, Vol.43, pp.689–693, 1998.

[2] R.H. Istepanian, G. Li, J. Wu and J. Chu, "Analysis of sensitivity measures of finite-precision digital controller structures with closed-loop stability bounds," *IEE Proc. Control Theory and Applications*, Vol.145, No.5, pp.472–478, 1998.

[3] D.S.K. Chan,"Constrained minimization of roundoff noise in fixed-point digital filters," in *Proc. ICASSP'79*, April 1979, pp.335–339.

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

[5] J. Yan and S.E. Salcudean, "Teleoperation controller design using optimization with application to motion-scaling," *IEEE Trans. Control Systems Technology*, Vol.4, No.3, pp.244–258, 1996.

[6] E.H.L. Aarts and J.H.M. Korst, *Simulated Annealing and Boltzmann Machines.* John Wiley and Sons, 1989.

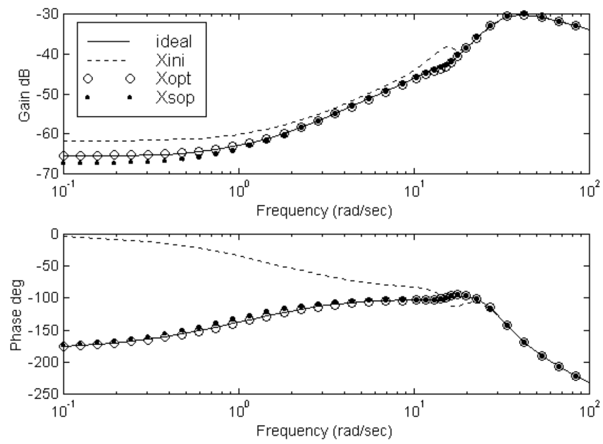

Figure 2: Frequency response plots for different realizations.