MTNS 1991 C60

COMPUTATIONALLY OPTIMAL IMPLEMENTATIONS VIA POLYNOMIAL OPERATORS

E.I. Verriest, G. Li, and M. Gevers

ABSTRACT

A new parametrization quadruple of a transfer function is suggested in terms of some generalized polynomial operators, for which the effective computational sensitivity is easily analysed. Optimal implementations are derived.

Key Words: Finite word length effects, parametrization.

1. INTRODUCTION

The use of the delta-operator form, was motivated in a recent text book [1]. It was argued that the δ form was more insightful, since the coefficients of a companion form realization in this δ form are close to the corresponding coefficients in the continuous model. The numerical properties of the implementation resulting from the delta operator representation are also noted to be superior. Of course, the δ-operator is not directly implementable in state space form, since it is not implemented by a causal structure, i.e. past data determining the present output by means of a sufficient statistic. The standard δ-operator form is a companion form corresponding to

$$\delta x_k = Ax_k + bu_k \tag{1}$$

$$y_k = cx_k \tag{2}$$

transformed (via $\delta = [z-1]/\Delta$) to a causally implementable form consisting of two parts: a (companion form) implementation of $(Ax_k+bu_k)=p_k$, and the causal construction $x_k+\Delta p_k$. In essence this construction boils down to replacing the "basis"-polynomials $z^n, z^{n-1}, \ldots, z, 1$ by the forms $\delta^n, \delta^{n-1}, \ldots, \delta, 1$ as function of z. In this paper, such a construction is extended to larger classes of polynomial operators, and optimal implementations are derived within this class.

2. POLYNOMIAL OPERATOR REPRESENTATIONS

The realization of a transfer function via the so-called canonical forms [2] are here generalized. We restrict our attention to the class of strictly proper transfer functions, whose denominator is monic.

$$H(z) = b(z)/a(z) = (b'S(z))\{z^n+a'S(z)\}^{-1} \tag{3}$$

where $a'=[a_1,\ldots,a_n]$, $b' = [b_1,\ldots,b_n]$, and $S(z)' = [z^{n-1},\ldots,z,1]$. The polynomials in the vector $[z^n, S(z)]$ play a role as indicators of a partial state. We shall now consider another polynomial representation of the transfer function, by selecting a different set of representation polynomials. So let T be the matrix transforming the polynomial vector $[z^n, S(z)']'$ to $[z_0(z), Z(z)']'$, where we degree of Z(z) be n-1. With these constraints, T is of the form

$$T = \begin{bmatrix} 1 & \tau' \\ 0 & T \end{bmatrix} \tag{5}$$

Since the coefficients of z^n in numerator and denominator are always respectively 0 and 1, they are redundant. The transformation matrix is fully parametrized by the n-vector τ and the n by n matrix T. It is not necessary that the submatrix T, and hence T be invertible, as was assumed by Li [3]; the transfer function $H(z) = [z+1]/[z^2+2z+3]$ is also represented, corresponding to a singular transformation, by $H(z) = ([0,1,0][z^2+2z+3,z+1,0]')/([1,0,0][z^2+2z+3,z+1,0]')$. A generalized transfer function is then specified in terms of *coefficients* (acting as coordinates) and *polynomials* (acting as frames). Let H_0 be the class of these polynomial operator parametrizations. A particular element will be denoted by the quadruple

$$H_o[(\alpha,\beta),(z_o(z),Z(z))] \in R^n \times R^n \times MR_n[z] \times (R_{n-1}[z])^n \tag{6}$$

where $MR_n[z]$ is the set of monic polynomials of order n in z. For the above example, the two generalized polynomial representations are therefore $H_o[((2,3)',(1,1)'),(z^2;z,1)]$ and the generalized transfer function $H_o[((0,0)',(1,0)'),(z^2+2z+3;z+1,0)]$.

Letting $Rat_n[z]$ denote the set of the strictly proper rational functions of z of degree n, the map H

$$H\colon H_o = R^{2n} \times MR_n[z] \times R_{n-1}[z]^n \longrightarrow Rat_n[z] \tag{7}$$

specified by the rational function $H = \beta' Z(z)/[z_o(z)+\alpha' Z(z)]$ in terms of the coordinates and the polynomial frame defines an *observable* in the terminology of [4]. Alternatively, one could specify the transformation T, i.e. the matrix T, together with the vector τ which generates $[z_o(z),\ Z(z)']$ from the "standard" $[z^n,\ S(z)']$. We denote this class of parametrizations by $H_1[(\alpha,\beta),(\tau,T)]$.

A class of restricted polynomial operators was considered in [3], for the case where T is restricted to be nonsingular. By relaxing the requirement that T be nonsingular, a new equivalence is defined in the class H_o of polynomial operator representations, for which the observable H(z) (the transfer function) is invariant.

Definition (of Equivalence): The polynomial operator representations $H_o[(\alpha_1,\beta_1),(z_{o1},\ Z_1)]$ and $H_o[(\alpha_2,\beta_2),(z_{o2},Z_2)]$ are equivalent (~) iff there exists vectors τ_i and matrices T_i, for i = 1,2 such that

$$z_{oi}(z) = z^n + \tau_i' S(z) \tag{8}$$

$$Z_i(z) = T_i S(z) \tag{9}$$

$$T_1'\alpha_1 + \tau_1 = T_2'\alpha_2 + \tau_2 \tag{10}$$

$$T_1'\beta_1 = T_2'\beta_2 \tag{11}$$

It is clear that the common vectors $a = T_i'\alpha_i + \tau_i$ and $b = T_i'\beta_i$ uniquely specify the denominator and numerator of the transfer function H(z). It is easily verified that the above is an equivalence relation. H is constant on the equivalence classes of H_o.

and Z will be demon-

3. GENERALIZED REACHABLE IMPLEMENTATIONS

We start from the polynomial operator parametrization $H_o[(\alpha,\beta),(z_o,Z)]$ as a right fraction description (rfd). The relation between input and output is then $y_t = \beta' Z(z)[z_o(z)+\alpha' Z(z)]^{-1}u(t)$. Via Kelvin's principle, this leads to a "natural" implementation, akin to the construction of the canonical realization. First this rfd separates into

$$y_t = \beta' Z(z)\zeta_t \tag{12}$$

$$[z_o(z)+\alpha' Z(z)]^{-1}u_t = \zeta_t \tag{13}$$

In turn (13) leads to

$$z_o(z)\zeta_t = u_t - \alpha_1 z_1(z)\zeta_t - \ldots - \alpha_n z_n(z)\zeta_t \tag{14}$$

where the polynomial $z_o(s)$ plays a privileged role, and will be referred to as the CORE-POLYNOMIAL. Note that it has the highest degree n. With this core-polynomial, a CORE-REALIZATION of

$$z_o(z)\zeta_t = p_t \tag{15}$$

is derived in a feed-out companion form [2], by implementing (15) as

$$z^n\zeta_t = p_t - (z_{o1}z^{n-1}\zeta_t + \ldots + z_{on}\zeta_t) \tag{16}$$

This "core of the core" is identified as the usual chain of n delay elements, ζ_t is the partial state, and $\chi_t = [z^{n-1},\ldots,z,1]'\zeta_t$ is the CORE STATE. It follows from (16) that the core-realization of (14) consists of a weighted feedback of the core-states χ_t superposed on p_t, and fed to the input of the core-core.

The state space representation of this feed-out core implementation follows:

$$\chi_{k+1} = A_c(\tau)\chi_k + b_c p_k \tag{17}$$

$$\zeta_t = [0,\ldots,0,1]\chi_t \tag{18}$$

where $A_c(\tau)$ is a top companion form matrix with characteristic polynomial $z^n+\tau' S(z)$, and $b_c = [1,0,\ldots,0]'$. Next, p_t is assembled by identifying the right hand sides of (14) and (15).

$$p_t = u_t - \alpha_1 z_1(s)\zeta_t - \ldots - \alpha_n z_n(s)\zeta_t \tag{19}$$

Define now the STATE of the implementation x_t by

$$[z_1(z),\ldots,z_n(z)]'\zeta_t = T\,[z^{n-1}(z),\ldots,z,1]'\zeta_t = T\chi_t = x_t \tag{20}$$

The core-input p_t in (19) is generated by a linear combination of input and realization states $p_t = u_t - (\alpha_1,\ldots,\alpha_n)x_t$ and finally, the system output follows from (12): $y_t = \beta' Z(z)\zeta_t = [\beta_1,\ldots,\beta_n]x_t$.

The implementation requires three inner products: $\langle\alpha,x\rangle$, $\langle\beta,x\rangle$ and $\langle\tau,\chi\rangle$, together with the matrix vector product $T\chi$. The latter should have as many "fixed" (i.e. zeros or ones) elements as possible, since these do not contribute any perturbation terms.

The implementation and its states can be interpreted as follows: The core consists of a bank of n filters, with transfer functions $z_1/z_o,\ldots,z_n/z_o$. All these are driven by p. Their outputs are the core-states. Feedback of the core-states is performed around the core with gains α, while the read-out of the core states is with the linear combiner β. Note that the actual number of independent filters in the core filter bank equals the rank of T. In particular, if T has rank one, then there is only a companion form realization of b/z_o, with scalar output feedback over α_1, and output scaling with β_1.

It was already pointed out that the δ-operator realizations can yield excellent roundoff noise performance. In fact, barring minor changes, all of the above implementations results can also be expressed in terms of the δ-operator rather than the shift operator z [5]. Some of our simulation results show again the numerical superiority of such implementations, however to illustrate the main ideas, the shift operator is used throughout the remainder of this paper.

4. SENSITIVITY OF THE IMPLEMENTATION.

The reason for distinguishing between realization and implementation is that the latter allows a more direct analysis of the sensitivity, as it involves directly the parameter values that are *implemented*, and not combinations thereoff. A singular system sensitivity analysis as in [6] is also not suitable because of the many fixed elements in the implementation. Conversely, changes in these parameters show up directly in the associated observable H.

Since H_1 is parametrized by α, β, τ, and T, the sensitivities of H, as functions of z, are obtained from the gradients

$$\partial H(z)/\partial\alpha = -TH(z)S(z)/a(z) \tag{21}$$

$$\partial H(z)/\partial\beta = TS(z)/a(z) \tag{22}$$

$$\partial H(z)/\partial T = [a(z)\beta - b(z)\alpha]S(z)'/a(z)^2 \tag{23}$$

$$\partial H(z)/\partial\tau = -H(z)S(z)'/a(z) \tag{24}$$

with the constraints $a = T'\alpha + \tau$ and $b = T'\beta$. The frequency dependent sensitivity measure is the sum of the H_2-norms of these quantities. For a parameter ρ this H_2^2 norm is

$$M_\rho = \|\partial H(\cdot)/\partial\rho\|_2^2 = (2\pi j)^{-1} \oint [\partial H/\partial\rho][\partial H/\partial\rho]^H z^{-1}\, dz \tag{25}$$

The contour integral is over the unit circle $|z|=1$, and the superscript H denotes the Hermitian conjugate. Noting that on the unit circle, $z^* = z^{-1}$, and that $S(z)'S(z^{-1}) = n$, one finds [3]

$$M_\alpha = \|\partial H(\cdot)/\partial\alpha\|_2^2 = \mathrm{Tr}\ (TW_\alpha T') \tag{26}$$

$$M_\beta = \|\partial H(\cdot)/\partial\beta\|_2^2 = \mathrm{Tr}\ (TW_\beta T') \tag{27}$$

$$M_T = \|\partial H(\cdot)/\partial T\|_2^2 = \mathrm{Tr}\ (c_1\beta\beta' - 2c_2\beta\alpha' + c_3\alpha\alpha') \tag{28}$$

$$M_\tau = \|\partial H(\cdot)/\partial\tau\|_2^2 = \mathrm{Tr}\ W_\alpha \tag{29}$$

The constants c_i, i=1,2,3 and the W_α and W_β are given by

$$c_1 = (2\pi j)^{-1} \oint n\ (za(z)a(z^{-1}))^{-1}\, dz \tag{30}$$

$$c_2 = \mathrm{Re}[\ (2\pi j)^{-1} \oint n\ H(z^{-1})\ (za(z)a(z^{-1}))^{-1}\, dz\ ] \tag{31}$$

$$c_3 = (2\pi j)^{-1} \oint n\ H(z)H(z^{-1})\ (za(z)a(z^{-1}))^{-1}\, dz \tag{32}$$

$$W_\alpha = (2\pi j)^{-1} \oint H(z)H(z^{-1})S(z)S(z^{-1})'(za(z)a(z^{-1}))^{-1}\, dz \tag{33}$$

$$W_\beta = (2\pi j)^{-1} \oint S(z)S(z^{-1})'(za(z)a(z^{-1}))^{-1}\, dz \qquad (34)$$

The latter is in fact the reachability gramian for the reachable realization. The sensitivity term M_τ is implementation independent, and therefore just adds a constant term to the overall sensitivity. The optimal implementation is therefore obtained by constrained minimization of the quantity $M = M_\alpha + M_\beta + M_T = \mathrm{Tr}\ TWT' + M_T$, and leads to a particular solution

$$T^* = \begin{bmatrix} [\Gamma/b'Wb]^{1/4} b' \\ \hline 0 \end{bmatrix}, \quad \Gamma = (c_1 c_3 - c_2^2)/c_3 \qquad (35)$$

$$\beta^* = [b'Wb/\Gamma]^{1/4} e_1 \qquad (36)$$

$$\alpha^* = ((c_2/c_3)\ [b'Wb/\Gamma]^{1/4}) e_1 \qquad (37)$$

$$\tau^* = a - (c_2/c_3) b \qquad (38)$$

The total optimal performance $M + M_\tau$ is

$$M_{opt} = 2[b'Wb]^{1/2}\Gamma^{1/2} + \mathrm{Tr}\ W_\alpha \qquad (39)$$

Although the sensitivity term with respect to τ is implementation independent, it may contribute largely to the overall sensitivity. However, the above sensitivity was based on an infinitessimal analysis. In digital computing, the perturbations are not infinitessimal but rather quantized. This has for effect that if for instance the parameters τ are choosen a priori in a perfectly implementable form, then the sensitivity with respect to this τ must vanish. This fact was not captured in our infinitessimal model. In view of this observation, one can make the analysis faithful to the computation environment by minimizing the performance measure M with respect to α, β, and T only (while keeping τ^o fixed). From the solution to the optimization problem, we know that τ close to τ^* is near optimal. Hence let τ^o be for instance an 8-bit FWL approximation to the optimal τ^*. With fixed τ^o, the optimization leads to the necessary condition

$$PWP = M_c \qquad (40)$$

where now

$$M_c = c_1 bb' - c_2(ba^{*\prime} + a^* b') + c_3 a^* a^{*\prime} \qquad (41)$$

which in general will have rank 2. The solution to this sub-optimal formulation does not necessarily require that a^* and b are parallel. Note that whereas analytically, the solution is suboptimal, for practical purposes, its performance will be superior, since the additional term M_τ now disappears, thus greatly reducing the overall sensitivity!

As indicated in our introduction, the performance of the implementation depends strongly on the choice of the core-realization. The δ-operator forms are numerically much better than the shift-operator forms. For this reason, we could just as well have considered the generalized δ-operators [5]. In fact some of the numerical simulations that we shall discuss below were exactly implemented in such a δ-form.

5. NUMERICAL VERIFICATION

For the realization in state space form, the performance was computed for the controller canonical form, and the balanced realization.

These were then contrasted to the implementations for fixed choices of τ, namely $\tau = 0$, a, τ^* and an exactly implementable finite wordlength (8 bits) version of τ. The latter are given for the δ- and the shift (z-)operator. For the δ-operator with 8-bit implementation of τ^*, a good coefficient dynamical range resulted as well. The latter implementation has also been simulated, and for an 8-bit word length, a performance exceeding that of the balanced state-space realization was observed.

6. CONCLUSIONS

1) A new implementation of systems, parametrized by a quadrapule (α,β,τ,T) was presented. This form allows the discrimination of exactly implementable, and non-exactly implementable operations, towards the coefficient sensitivity problem. The sensitivity of the transfer function with respect to the free parameters is optimized, and it was found that the optimal structure consists of a scalar feedback and output scaling of some implemented core system. The realization of the core itself, greatly influences the overall sensitivity. For the shift operator, we have taken this to be the companion form, but for a δ-operator implementation a much better performance was found [5].

2) If T is restricted to be nonsingular, then an extremum does not exist. This indicates that the extremum actually sits on the boundary, where T is singular. Hence only asymptotically optimal solutions can be obtained in this case.

REFERENCES

[1] Middleton, R.H. and G.C. Goodwin (1990). *Digital Control and Estimation*, Prentice Hall, Englewood Cliffs.

[2] Kailath, T. (1980). *Linear Systems*, Prentice Hall.

[3] Li, G. (1990). *Finite Precision Aspects in the Parametrization of Control, Estimation, and Filtering Problems*, Ph.D. Dissertation, Université Catholique de Louvain, Belgium.

[4] Verriest, E.I., and W.S. Gray, (1988). *Robust design problems: A geometric approach*, in *Linear Circuits, Systems and Signal Processing: Theory and Application*, C.I. Byrnes, C.F. Martin and R.E. Saeks, eds. Elsevier, North-Holland.

[5] Verriest, E.I., G. Li, M. Gevers. *Computationally Efficient Implementations via Polynomial Operators*, in preparation.

[6] Gray, W.S., E.I. Verriest, and F.L. Lewis (1990). A Hankel matrix approach to singular system realization theory. *Proc. 29th IEEE CDC*, Honolulu, HI.

Acknowledgement: The results presented in this paper have been obtained within the framework of the Belgian Program on Concerted Research Actions and on Interuniversity Attraction Poles initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility rests with its authors.