Alternative Discrete-Time Operators: An Algorithm for Optimal Selection of Parameters

Andrew D. Back, Bill G. Horne, Ah Chung Tsoi, C. Lee Giles

Abstract—**In this note, we consider the issue of parameter sensitivity in models based on alternative discrete time operators (ADTOs). A generic first order ADTO is proposed which encompasses all the known first order ADTOs introduced so far in the literature. New bounds on the operator parameters are derived, and a new algorithm is given for optimally selecting the parameters to give minimum parameter senstivity.**

Keywords—**Digital filters, low sensitivity, delta operator, shift operator, gamma operator.**

I. Introduction

An important aspect in the development of linear models, are practical properties such as parameter convergence, model robustness, parameter sensitivity, and parsimony in representations.

In this note we consider the problem of parameter sensitivity which occurs when using a finite wordlength representation. Suppose we have a hardware implementation using two's complement arithmetic. Then any real valued number a can be represented in B bits as $a = k\left(-b_0 + \sum_{i=1}^{B-1} b_i 2^i\right) + e$ where k is some scaling factor, b_0 is the sign bit, and e is the error in the representation [1]. The question of interest, is what effect does e play in the properties of a given model for some finite B? When fixed point arithmetic is used, the issue of quantization of the signals and of the model parameters needs to be considered.

A variety of methods have been proposed to deal with the problem of quantization in linear systems. It is well known that for any linear filter transfer function there exist many possible structural realizations, all of them related through a similarity transformation. The problem of determining an optimal filter realization which minimizes the round-off noise has been solved by Mullis and Roberts [2]. Theile solved the problem of determining an optimal similarity transformation to reduce parameter sensitivity [3, 4]. It has been shown by Jackson [5] that a low noise structure can also have low parameter sensitivity properties provided certain constraints are met. Various filter realizations with low round-off noise and low parameter sensitivity have been proposed in [6–8].

Another widely used approach to produce models with low parameter sensitivity is the method of shift operator replacement. Such operators have become known as *alternative discrete time operators*. Alternative discrete time operators (ADTOs) have been proposed in various forms for improving the numerical performance of control and system identification models, and for digital filters [9–13]. A number of ADTOs have been proposed in the literature. Some of most common examples are:

1. *Delta operator:* $\delta = \frac{z-1}{\Delta}$ where Δ is the sampling rate. This operator was introduced for digital filters in [9] to replace the shift operator. It has since been shown to have better performance than the shift operator in direct form digital filters [14] and in other applications [10, 15].

2. *Gamma operator:* $\gamma = \frac{z-(1-c)}{c}$, where c is an adjustable parameter. This operator was introduced in [16] as an efficient basis of convolutional memory representation.

3. *Rho operator:* $\rho = \frac{z-(1-c_1\Delta)}{c_2\Delta}$, where c_1, c_2 are adjustable parameters. This operator was introduced in the context of robust adaptive control [12]. It is stably invertible, thus allowing the derivation of robust estimation algorithms.

4. *Other structures:* Bilinear transform operators were proposed independently in [11, 17], a second order γ operator was introduced in [18] and a number of other possibilities are mentioned in [10]. The Laguerre filter [13] also gives low sensitivity behaviour. The Kautz filter [19] is an extension of the Laguerre filter, but allows for complex poles.

While these operators have been of particular interest in the area of high speed sampling [10, 11, 15] and provide low parameter sensitivity models, there appears to have been little effort given to determine the optimal selection of the operator parameters. One reason for this is that the most commonly used structure is the δ operator which has no adjustable parameters. Adaptive algorithms have been proposed for the γ operator in the context of minimizing mean square output error, but not for producing low sensitivity models. Other work performed in this area has been directed towards finding appropriate model parameters which are based on ADTOs [11, 15].

In this paper we consider a generic first order ADTO which is described in Section II. We give bounds on the operator parameters and an algorithm for the optimal selection of parameters to obtain minimum parameter sensitivity in a linear filter. A simple numerical example is given to illustrate the effectiveness of the algorithm. Conclusions are given in Section III.

II. Analysis and Selection of Operator Parameters

A. A Generic First Order ADTO: the ν operator

For the purposes of this paper we will consider a generic first order operator given by

$$\nu = \frac{z - c_1}{c_2} \tag{1}$$

where c_1, c_2 are constants. The generic operator may represent each of the previously discussed types: δ, γ and ρ operators by appropriate choice of the pararameters. The operator is stable for $0 < |c_1| < 1$.

In this section, we will present some new results on the bounds on c_1 and c_2 and the *optimal* selection of these parameters to obtain minimum parameter sensitivity models. The derivation of these results is not difficult, though they give properties of the ν operator which have not appeared previously in the literature[1].

Andrew Back is with the Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako–shi, Saitama 351-0198 Japan. Email: back@brain.riken.go.jp. Bill Horne is with AADM Inc., 9 Pace Farm Road, Califon, NJ 07830,USA. Ah Chung Tsoi is with the Faculty of Informatics, University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia. C. Lee Giles is with NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA and also with the Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742 USA. ADB acknowledges financial support from the Australian Research Council and the Brain Science Institute, RIKEN, Japan. ACT acknowledges partial support from the Australian Research Council.

[1] The lower bound of c_2 is a generalization of the results derived for the shift operator which appeared in [11]. The results on optimal parameter sensitivity are new, as far as we are aware.

B. Bounds on the Selection of Operator Parameters

For a polynomial $H(z) = \sum_{i=0}^{n} b_i z^{-i} = z^{-n} \prod_{i=1}^{n}(z - \lambda_i)$, where λ_i is the ith root (ie. a zero of the filter), with parameter sensitivity $S_{ij} = -\frac{\partial \lambda_i}{\partial b_j}$, we seek another rational function $H(\nu) = \sum_{i=0}^{n} b'_i \nu^{-i} = \nu^{-n} \prod_{i=1}^{n}(\nu - \lambda'_i)$, $\lambda'_i = \frac{\lambda_i - c_1}{c_2}$; and $S'_{ij} = -\frac{\partial \lambda'_i}{\partial b'_j}$, such that $S'_{ij} \leq S_{ij}, \forall i, j$ and $H(z) = H(\nu)$.

Following a similar approach to the δ operator case [10, 11], the parameter sensitivity is found as

$$\begin{aligned} S'_{ij} &= \frac{(\lambda'_i)^{n-j}}{\prod_{k \neq i}(\lambda'_i - \lambda'_k)} \\ &= c_2^{j-1} \frac{(\lambda_i - c_1)^{n-j}}{\prod_{k \neq i}(\lambda_i - \lambda_k)} \end{aligned} \tag{2}$$

A lower bound on c_2 may be found as follows. For polynomials $B(z) = \sum_{i=0}^{n} b_i z^i$ and $B(\nu) = \sum_{i=0}^{n} b'_i \nu^i$ it can be shown that the parameters b_j and b'_j are related by the following formula:

$$b'_i = c_2^i \sum_{j=i}^{n} b_j \begin{pmatrix} j \\ i \end{pmatrix} c_1^{j-i} \tag{3}$$

Hence, it is clear that if $c_2 < 1$, then $b'_n \to 0$ as n becomes large. Consider the leading coefficient, $b'_n = b_n c_2^n$. This term vanishes as n becomes large which implies that for FWL implementations of *polynomials*, the coefficients may end up getting truncated to zero if c_2 is too small. A lower bound on c_2 is given by

$$c_2 \geq 2^{-\frac{(b-1)}{n}} \tag{4}$$

where b is the wordlength in bits used to represent c_2.

This lower bound can be used to select either a minimum c_2 value for a given number of bits representation, or a minimum word length for a given c_2 value. There is not any practical upper bound on c_2, since if $\max_i S'_{ij}(c_1, c_2)$ occurs for $j = 0$, then c_2 should be as large as possible. Note that from these constraints, the operator ν will always be a *low pass* filter.

C. Optimal Operator Sensitivity

The next question we seek to answer, is how to select c_1 and c_2 to *optimize* this improvement in sensitivity for a given FWL implementation[2].

From (2) we know the sensitivity of the roots of $H(\nu)$, is a function of the coefficients of the corresponding polynomial in c_1, c_2. Note that the sensitivity $S'_{ij}(c_1, c_2)$ consists of a matrix of functions, one corresponding to each root and coefficient. The approach we adopt, is to seek to find the overall operator parameters c_1, c_2 which will simultaneously minimize the largest of all of the sensitivities given by

$$E = \max_{i,j} \left| S'_{ij}(c_1, c_2) \right| \tag{5}$$

To minimize (5) with respect to c_1, it is sufficient to simultaneously minimize the maximum of the partial sensitivity functions f_i defined as

$$f_i(c_1) = \frac{|\lambda_i - c_1|^n}{\prod_{k \neq i} |\lambda_i - \lambda_k|} \quad i = 1, ..., n \tag{6}$$

To minimize the maximum of (6) over $i = 1, ..., n$ we proceed as follows. First, each f_i is a simple function concave upwards everywhere, and has a unique minimum at $c_{1i}^* = \Re\{\lambda_i\}$, where $\Re\{\lambda\}$ indicates the real part of λ. Since $\max_i f_i$ may be formed piecewise from various f_i $i = 1..n$, then it is also concave upwards. It follows that the required value of c_1^* which will minimize $\max_i f_i$ may occur at one of the c_{1i}^*, or where two of the functions, denoted by f_{i^*} and f_{j^*}, intersect. The required intersecting curves f_{i^*} and f_{j^*} are those which define, piecewise, $\max_i f_i(c_1)$, $i = 1, ..., n$.

The method we use to find the functions f_{i^*} and f_{j^*} is similar to that used in convex hull algorithms [20]. Firstly we find the *minimums* of each individual f_i $i = 1, ..., n$, and the intersection of each possible pair of functions. We then perform a search to find the *maximum* value of the function or functions (intersecting case). We then choose c_1^* as the value which will minimize $\max_i f_i$.

The intersection of f_{i^*} and f_{j^*} occurs when

$$\frac{|\lambda_{i^*} - c_1|^n}{\prod_{k \neq i^*} |\lambda_{i^*} - \lambda_k|} = \frac{|\lambda_{j^*} - c_1|^n}{\prod_{k \neq j^*} |\lambda_{j^*} - \lambda_k|} \tag{7}$$

Defining $d_j = \left[\prod_{k \neq j} |\lambda_j - \lambda_k|\right]^{2/n}$, we may rewrite (7) as a quadratic equation in the unknown c_1,

$$a' c_1^2 + b' c_1 + c' = 0 \quad , \tag{8}$$

where $a' = \frac{1}{d_{i^*}} - \frac{1}{d_{j^*}}$, $b' = \frac{2\Re\{\lambda_{j^*}\}}{d_{j^*}} - \frac{2\Re\{\lambda_{i^*}\}}{d_{i^*}}$ and $c' = \frac{|\lambda_{i^*}|^2}{d_{i^*}} - \frac{|\lambda_{j^*}|^2}{d_{j^*}}$. The roots of this quadratic function may be real, real and repeated, or complex and give the intersection point.

If the roots of (8) are real and unique, then f_{i^*} and f_{j^*} will intersect at two points. One of those points will be the required minimum, corresponding to c_1^* (see Fig. 1(a)). If the roots of (8) are repeated, then the curves intersect at only one point, giving c_1^* directly. If the roots are complex, the curves f_{i^*} and f_{j^*} do not intersect(see Fig. 1(b)). In this case, the optimal c_1^* value can be found simply at the minimum of $\max(f_{i^*}, f_{j^*})$. This point can be obtained by taking the derivative of $\max(f_{i^*}, f_{j^*})$ with respect to c_1. A simple test for such complex roots is that $\lambda_{i^*}^2 + \lambda_{j^*}^2 < 2\lambda_{i^*}\lambda_{j^*}$.

The method for determining the true values i^* and j^* is shown in shown in Table I. Lack of space does not permit a detailed description, however the procedure essentially compares successive c_1^* values until convergence to find the min-max solution on the convex hull.

Once we have c_1^*, we then proceed to find c_2^*. For any given value of c_1, it is possible to choose c_2 such that $S'_{ij}(c_1^*, c_2)$ is a maximum for either $j = 0$ or $j = n$. Using (2), substituting $j = n$, we have

$$\max_i |S'_{in}(c_1^*, c_2)| = c_2^{n-1} \max_i g_i \tag{9}$$

[2] If infinite precision was possible, we could improve the sensitivity to some arbitrary value by simply making c_2 as small as we like.

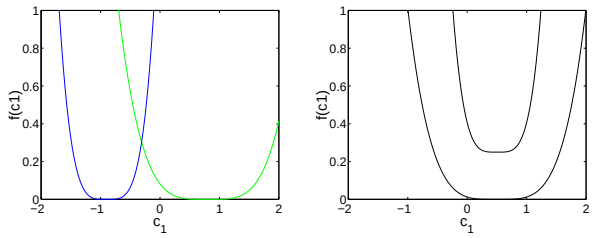

Fig. 1. (a) Intersection of f_{i^*} and f_{j^*} for the case where the roots of the quadratic equation are real. (b) For the case where the roots are complex, the curves do not intersect.

where $g_i = \frac{1}{\prod_{k \neq i} |\lambda_i - \lambda_k|}$ and similarly substituting $j = 0$ in (2) we have,

$$\max_i |S'_{i0}(c_1^*, c_2)| \quad = \quad c_2^{-1} \max_i f_i\,(c_1^*) \tag{10}$$

Note that (9) is an increasing function with c_2, while (10) is a decreasing function with c_2. Since $\max_i g_i$ and $\max_i f_i\,(c_1)$ are both positive reals, c_2^* can be found at the intersection of (9) and (10) which minimizes $\max_{ij} \left|S'_{ij}(c_1^*, c_2)\right|$. Hence c_2^* can be found by solving for the equality of (9) and (10) to give

$$c_2 \quad = \quad \left[\frac{\max_i f_i(c_1)}{\max_i g_i}\right]^{1/n} \tag{11}$$

The algorithm is summarized in Table I. To illustrate the ideas contained in the paper we give a numerical example below.

Example: Selection of parameters for a linear FIR model

Consider an FIR model given by:

$$\begin{aligned} G(z) \quad = \quad & 1.0 - 7.58z^{-1} + 24.3681z^{-2} - 42.4387z^{-3} \\ & -23.1061z^{-4} + 5.4960z^{-5} \end{aligned} \tag{12}$$

The maximum parameter sensitivity of each of this filter is $\max_{i,j} S_{ij} = 1829$. The model is implemented using ADTOs and applying the proposed algorithm, we obtain parameter values $c_1 = 1.0$ and $c_2 = 0.705$. The maximum parameter sensitivity now obtained is $\max_{i,j} S'_{ij} = 21$, an improvement of more than 87 times over the usual shift operator case.

To see the effect this has on the actual performance of the filter, we simulate the models with 8 bit wordlength. It can be easily observed in Fig. 2 that the $H(z)$ model is subject to significant changes in the zero positions due to the FWL implementation. The difference between the true zeros (shown as 'o') and the zeros due to coefficient rounding (shown as '*') can be clearly seen. In Fig. 2(b) the performance of the $H(\nu)$ model is shown. Here there is no significant change in the zero positions due to coefficient rounding. From this example[3], it is evident that the ν operator model exhibits significantly better performance than the shift operator model.

[3] We have conducted a number of experiments using different models and they show similar performance improvements.

TABLE I

ALGORITHM FOR DETERMINING OPTIMAL ν OPERATOR PARAMETERS.

For $B(z)$, sort λ_i $i = 1 \ldots n$: $\Re\{\lambda_1\} < \cdots < \Re\{\lambda_n\}$
let $i^* = 1, j^* = n$
// Determine true i^*, j^* to find optimal c_1^*.
do
  for $i = 2, .., n-1$
    function c'_{1i} = compute c1(i, j^*)
    if $f_i(c'_{1i}) > f_{i^*}(c_1^*)$
      $i^* := i$ **fi**
  end
  for $j = n-1, .., 2$
    function c'_{1j} = compute c1(i^*, j)
    if $f_j(c'_{1j}) > f_{i^*}(c_1^*)$
      $j^* := j$ **fi**
  end
until convergence of c_1^*
if c_1^* > 1.0 **then** c_1^* = 1.0
$c_2^* = \left(\frac{\max_i f_i(c_1)}{\max_j g_j}\right)^{1/n}$.

function compute $c_1(i, j)$
// For any f_i and f_j, find the min-max solution point c_1:.
  if $\lambda_i^2 + \lambda_j^2 < 2\lambda_i\lambda_j$
// Test for nonintersecting f_i and f_j.
    // Roots of (8) are complex.
    For any c_1, find $\tilde{i} = \arg\max_{i=i,j} f_i\,(c_1)$.
    $c_1 = \Re\{\lambda_{\tilde{i}}\}$.
  else
    // Roots of (8) are real.
    For i and j, solve (8) for c_1.
  fi

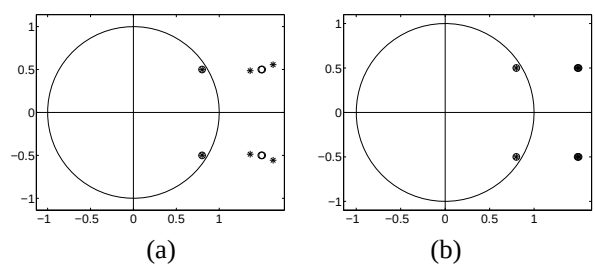

Fig. 2. Example results showing the zero positions for: (a) $H(z)$ and (b) $H(\nu)$, due to coefficient rounding using an 8 bit wordlength. In each case the true zeros are shown as circles and the perturbed zeros due to coefficient rounding are shown using the * symbol. The lower parameter sensitivity of the ν operator model gives significantly reduced change in the zeros when the coefficients are rounded.

III. CONCLUSIONS

In this paper we briefly reviewed a number of discrete time operators which may be used in place of the usual shift operator in digital filters. These operators provide a means of giving significantly improved performance under conditions of parameter perturbation. As a means of dealing with the various ADTOs,

we considered a generic first order operator, termed the ν operator. We derived an algorithm for choosing optimal parameters in terms of reducing the maximum parameter sensivity. The effective performance of the algorithm has been shown in experiments, and we show a typical result in this paper.

It is apparent that alternative discrete-time operators offer a useful approach for deriving low sensitivity linear models which can give significantly improved numerical accuracy when compared with conventional shift operator structures.

REFERENCES

[1] R.A. Roberts and C.T. Mullis, *Digital Signal Processing*, Addison-Wesley, 1987.

[2] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff fixed point digital filters," *IEEE Trans. Circuits, Syst.*, pp. 551–562, 1976.

[3] L. Thiele, "Design of sensitivity and roundoff noise optimal state-space discrete systems," *Int. J. Circuit Theory Appl.*, vol. 12, pp. 39–46, 1984.

[4] L. Thiele, "On the sensitivity of linear state-space systems," *IEEE Trans. Circuits, Syst.*, vol. CAS-33, pp. 502–510, 1986.

[5] L.B. Jackson, *Digital Filters and Signal Processing*, Kluwer Academic Publishers, Norwell,MA, 1996.

[6] L.B. Jackson, A.G. Lindgren, and Y. Kin, "Optimal synthesis of second order state space structures for digital filters," *IEEE Trans. Circuits, Syst.*, vol. CAS-26, no. 3, pp. 149–153, 1979.

[7] C.W. Barnes, "Roundoff noise and overflow in normal digital filtering," *IEEE Trans. Circuits, Syst.*, vol. CAS-26, no. 3, pp. 154–159, 1979.

[8] W. L. Mills, C. T. Mullis, and R. A. Roberts, "Low roundoff noise and normal realizations of fixed point IIR digital filters," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 4, pp. 893, 1981.

[9] R.C. Agarwal and C.S. Burrus, "New recursive digital filter structures having very low sensitivity and roundoff noise," *IEEE Trans. Circuits, Syst.*, vol. CAS-22, no. 12, pp. 921–927, 1975.

[10] G.C. Goodwin, R.H. Middleton, and H.V. Poor, "High-speed digital signal processing and control," *Proc. IEEE*, vol. 80, no. 2, pp. 240–259, 1992.

[11] M. Gevers and G. Li, *Parameterizations in control, estimation and filtering problems: accuracy aspects*, Springer-Verlag, London, 1993.

[12] M. Palaniswami, "Digital estimation and control with a new discrete time operator," in *Proc. 30th IEEE Conf. Decision and Control*, New York, NY, 1991, IEEE Press, pp. 1631–1632.

[13] B. Wahlberg, "System identification using laguerre models," *IEEE Trans. Automat. Control*, vol. 36, pp. 551–562, 1991.

[14] D. Williamson, "Delay replacement in direct form structures," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 4, pp. 453–460, Apr. 1988.

[15] H. Fan and Q. Li, "A δ-operator recursive gradient algorithm for adaptive signal processing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Proc.* IEEE Press, 1993, vol. III, pp. 492–495.

[16] B. de Vries and J.C. Principe, "A theory for neural networks with time delays," in *Advances in Neural Information Processing Systems*, R.P. Lippman, J.E. Moody, and D. S. Touretzky, Eds., San Mateo, CA, 1991, Morgan Kaufmann, vol. 3, pp. 162–168.

[17] A.D. Back and A.C. Tsoi, "A comparison of discrete-time operator models for nonlinear system identification," in *Advances in Neural Information Processing Systems*, G. Tesauro, D. S. Touretzky, and T. K. Leen, Eds., Cambridge, MA, 1995, The MIT Press, vol. 7, pp. 883–890.

[18] T. Oliveira e Silva, P.G. de Oliveira, J.C. Principe, and B. de Vries, "Generalized feedforward filters with complex poles," in *Proc. of the 1992 IEEE Workshop Neural Networks for Signal Processing 2 (NNSP92)*, S.Y. Kung, F. Fallside, J. Aa. Sorenson, and C.A. Kamm, Eds., Piscataway, NJ, 1992, IEEE Press, pp. 503–510.

[19] P. Lindskog and B. Wahlberg, "Application of Kautz models in system identification," in *Proc. 12th Triennial World Congress*. IFAC, 1993, pp. 41–44.

[20] J.R. Smart, *Modern Geometries*, Brooks/Cole, New York, 1998.