

# The Choice between Delta and Shift Operators for Low-Precision Data Representation

Denis N. Butusov, Timur I. Karimov,  
Dmitry I. Kaplun, Artur I. Karimov

Saint Petersburg Electrotechnical University "LETI"  
St. Petersburg, Russia  
butusovdn@mail.ru

Yennun Huang, Szu-Chuang Li

CITI, Academia Sinica  
Taipei City, Taiwan  
yennunhuang@citi.sinica.edu.tw

**Abstract**—Low-precision data types for embedded applications reduce the power consumption and enhance the price-performance ratio. Inconsistence between the specified accuracy of a designed filter or controller and an imprecise data type can be overcome using the  $\delta$ -operator, an alternative to the traditional discrete-time  $z$ -operator. Though in many cases it significantly increases accuracy, sometimes it shows no advantage over the shift operator. So the problem of choice between delta and shift operator arises. Therefore, a study on  $\delta$ -operator applicability bounds is needed to solve this problem and provide  $\delta$ -operator efficient practical use. In this paper we introduce a concept of the  $\delta$ -operator applicability criterion. The discrete system implementation technique with discrete-time operator choice is given for the low-precision machine arithmetic.

## I. INTRODUCTION

The standard technique for the linear dynamical system hardware representation is the shift operator  $z$ . The design of  $z$ -systems is based on well-known discretization methods like bilinear transform or zero-pole matching and is fully formalized now. However, the use of low-precision arithmetic results in high round-off errors and may strongly affect the performance of the designed system. To increase the accuracy an engineer has to use longer data types. This lowers some other valuable properties of the system: the power consumption, the speed of calculations and the overall system cost [1]. If the chosen hardware is FPGA, system will require more LUT's and DSP blocks. If the system is based on a microprocessor a more expensive and less power-efficient chip will be needed.

The  $\delta$ -operator was introduced [2] as an alternative to the shift operator. Zeros and poles of the  $\delta$ -operator discrete-time system behave differently from those in the  $z$ -operator system and do not converge to 1 or 0 as the sample period decreases. Experiments have shown that the  $\delta$ -operator may significantly enhance digital systems parameters. For example the paper [3] considers an implementation of an optimal sliding-mode control system with the  $\delta$ -operator. The resulting system much rarely suffers from self-oscillations than the shift operator based system. A study [4] compares two implementations of PID-controllers, based on  $\delta$  и  $z$ -operators respectively. It is shown that the difference equation coefficients for the  $\delta$ -system are less affected by machine word length restrictions.

Paper [5] describes the  $\delta$ -operator LQR controller implementation. Custom data type with only five mantissa bits provided an accuracy compared to the shift operator-based LQR controller with the IEEE single floating point datatype.

Though papers dedicated to  $\delta$ -operator shows incessant interest to this mathematical tool over the years, the question about  $\delta$ -operator applicability limits still has no exact answer. The problem is that  $\delta$ -operator implementation requires some complication of the system but not every system converted from  $z$  to  $\delta$  domain shows the significant increase of accuracy.

Known recommendations on the choice between  $z$  and  $\delta$ -operator are unsatisfactory fuzzy. In the most of studies, e.g. in papers [6], [7], the  $\delta$ -operator is introduced without any specific methodic of application. Some other publications state that the operation frequency of the system should be considerably lower than the Nyquist frequency for the chosen sample rate. Early papers are full of optimism and predict total change of discretization operator paradigm [8] that did not happen in fact. Thus, the discrete operator choice technique is important to be developed. A technique for some special cases of dynamical systems was considered in early paper [9]. The present article generalizes some obtained results and presents a set of solutions for a discrete operator choice problem. Section II introduces the general concept of delta operator preference criterion. Section III reviews the roundoff error influence on discrete filter behavior. Section IV discusses the performance of different discrete filter types with normalized frequency. Section V presents an enhanced algorithm for discrete operator selection with an improved choice criterion. Finally, conclusions are presented in Section VI.

## II. THE CONCEPT OF DELTA OPERATOR PREFERENCE CRITERION

Definition of the  $\delta$ -operator is given in [10], [11]. Briefly,  $z = e^{sT_s}$  determines the  $z$ -operator through Laplace operator  $s$ , then  $\delta = \frac{z-1}{\Delta}$  defines  $\delta$ -operator through  $z$ -operator. Symbol  $\Delta$  denotes a certain coefficient that can be almost arbitrary number. The prime idea was to establish  $\Delta = T_s$  i.e. sample period [10], later it was shown in [12] that the choice  $\Delta = 2^{-n}$ ,  $n \in \mathbb{N}$  is often more feasible for hardware

implementation. To simplify the computations and for the floating-point implementation the value  $\Delta = 1$  is also acceptable.

Conversion from z-domain to  $\delta$ -domain can be performed through the algebraic transformation of coefficients using the general formula

$$\beta_k = \frac{1}{\Delta^k} \left( b_k \binom{n-k}{n-k} + b_{k-1} \binom{n-k+1}{n-k} + \dots + b_0 \binom{n}{n-k} \right)$$

for the nominator and

$$\alpha_k = \frac{1}{\Delta^k} \left( a_k \binom{n-k}{n-k} + a_{k-1} \binom{n-k+1}{n-k} + \dots + 1 \cdot \binom{n}{n-k} \right)$$

for the denominator. Here  $\binom{l}{m}$  denotes binomial

coefficients:

$$\binom{l}{m} = \frac{l!}{m!(l-m)!},$$

$a_i, \alpha_i$  — denominator coefficients of z- and  $\delta$ -system respectively,  $b_i, \beta_i$  — numerator coefficients,  $n$  — system (section) order.

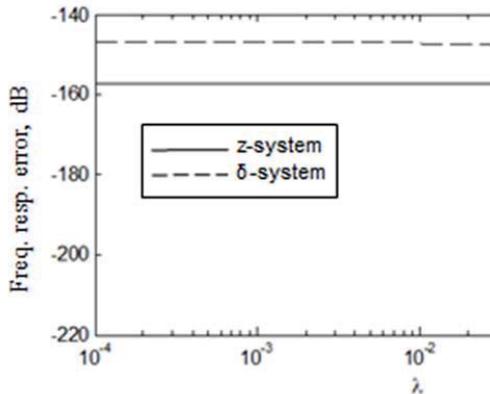


Fig. 1. Frequency response errors plots for z-operator and  $\delta$ -operator

The given example clearly shows that operator preference criterion depends on the normalized filter frequency. According to this observation, the generalized discrete operator choice algorithm is:

1) Check the condition  $\lambda_{op} \ll 1$  (operation frequency is considerably lower than the Nyquist frequency). If it is true, there may be a reason for  $\delta$ -operator application.

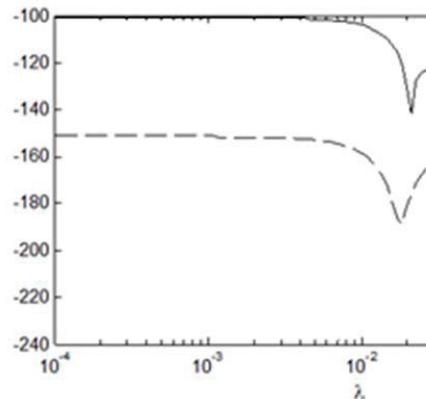
2) Find the value  $S = \|E_z\| / \|E_\delta\|$  where  $\|\cdot\|$  denotes a certain norm and  $E_z, E_\delta$  — frequency response error estimations. If  $S \geq S^*$ , where  $S^*$  is specified level, the

The performance of two operators is illustrated by the following example. Assume an embedded hardware supporting single precision floating point numbers operations. Synthesize 2<sup>nd</sup> order low-pass elliptic filter with the following specs: passband ripple  $10^{-4}$  dB, stopband attenuation 40 dB, normalized passband edge  $\lambda_s = 0,03$  (case 1) and  $\lambda_s = 0,002$  (case 2). Frequency response errors for both operators compared to the double precision counterpart are plotted in Fig. 1.

$$H_1(z) = \frac{0.1267(z^2 + 1.835z + 1)}{z^2 - 0.8055z + 0.2915} \Leftrightarrow H_1(\delta) = \frac{0.1267(z^2 + 3.835z + 3.835)}{z^2 + 1.1945z + 0.486} \Big|_{\Delta=1}, \quad (1)$$

$$H_2(z) = \frac{0.0105(z^2 - 1.627z + 1)}{z^2 - 1.91z + 0.9139} \Leftrightarrow H_2(\delta) = \frac{0.0105(z^2 + 0.373z + 0.373)}{z^2 + 0.09z + 0.0039} \Big|_{\Delta=1}. \quad (2)$$

In both cases mantissa has the length of 23 bits, but this is not sufficient for z-operator to show the same performance in the case 2 as in the case 1 while  $\delta$ -system shows no loss of accuracy. Also, it is obvious that  $\delta$ -operator is unnecessary in the case 1. The question is: under what conditions the  $\delta$ -operator is more precise than z?



$\delta$ -operator is preferable for this system. Further we assume  $S^* = 1$ , which corresponds to the  $\|E_z\| = \|E_\delta\|$  case.

Determination of the  $\delta$ -operator preference regions allows to roughly estimate the  $\delta$ -operator applicability for a certain discrete system at the initial design stage and can be useful in engineering. Moreover, from the mathematical point of view, this issue is of fundamental importance. The investigated regions provide a quantitative measure for the number of systems which are represented by means of the  $\delta$ -operator more precisely comparing to the shift operator. This will clarify which cases of successful  $\delta$ -operator application are a kind of engineering luck and which are reasonable. In paper [9] the given criterion was formulated in the following way.

Consider a discrete system with the state space matrix  $A$ . It has better accuracy in  $\delta$ -domain with 97% probability when eigenvalues of  $A$  lie inside the  $P_\delta$  region. On  $z$ -plane, the region  $P_\delta$  has approximately elliptic shape with the center in the origin of coordinates and axes  $(0.66, 1.15)$  — see Fig. 2. If the system has real eigenvalues then its preference region is the segment  $(-0.66, 0)$  on the real axis.

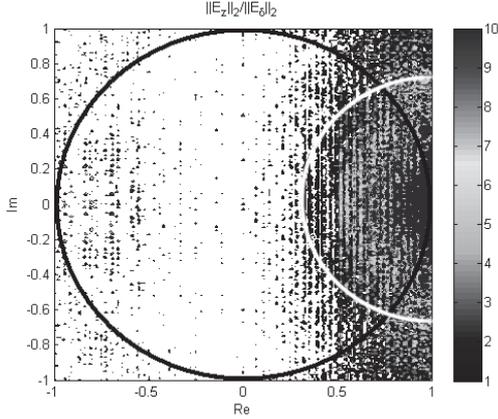


Fig. 2. Delta operator preference region outlined with the white border on  $z$ -plane. The black border outlines the unit circle

However, described discrete operator choice criterion is not very convenient in practice especially when implementing systems described as transfer functions. Also paper [9] does not answer the question how much the desired criterion is influenced by selected data type.

Following sections study the roundoff error with respect to the data and discrete operator type and define the condition  $\lambda_{op} \ll 1$  more strictly.

### III. THE ROUND OFF ERROR IN FILTER COEFFICIENTS

A computer representation error  $E_r$  of real number  $r$  can be calculated as

$$E_r(r) = r - [r] = \text{mod}\left(r + \frac{\varepsilon}{2}, \varepsilon\right) - \frac{\varepsilon}{2}, \quad (3)$$

where  $\varepsilon$  is the computer relative accuracy for number  $r$  or smallest representable number, and square brackets denote the computer number. The main idea of formula (3) is that roundoff error appears when «superfluous» bits are truncated to fit mantissa length with  $\text{mod}(\cdot, \varepsilon)$  function. The maximal error  $E_r = r - [r]$  equals

$$\max_r E_r = \frac{\varepsilon}{2},$$

because  $\text{mod}(\cdot, \varepsilon)$  gives values on the interval  $[0; \varepsilon)$ . Notice

that the error is equal to 0 in nodal points  $r + \frac{\varepsilon}{2}$ . The

described form of error can be illustrated through experimental data — see Fig. 3. For the given specification  $r \in [1; 1 + 3\varepsilon]$ ,

$\varepsilon = 2^{-8}$  (8 bits for mantissa) values were taken.

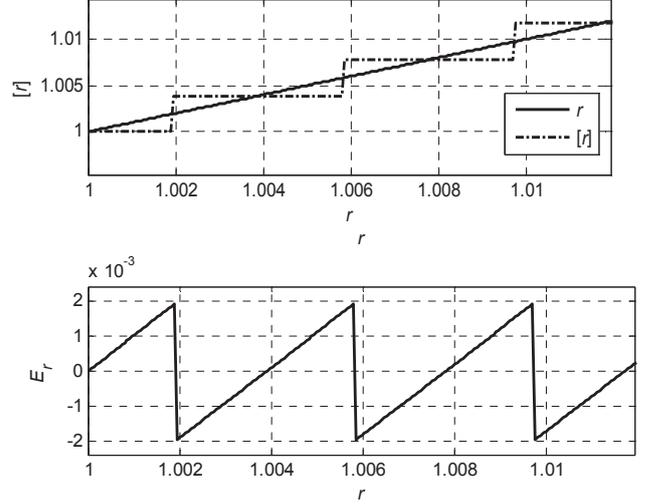


Fig. 3. Experimental estimation of a truncation error

This experimental study shows that there is no dependence between mantissa length and discrete operator preference criterion. Strict proof is given below.

Consider the following lemma.

**Frequency response error bounds lemma:** denote  $H(\lambda) = \frac{b(\lambda)}{a(\lambda)} = \frac{b}{a}$ ,  $\varepsilon$  — relative accuracy for number 0.5. Let  $\lim_{\lambda \rightarrow 0} |H(\lambda)| \neq 0$ , as well as nominator and denominator  $\lim_{\lambda \rightarrow 0} E_b \neq 0$ ,  $\lim_{\lambda \rightarrow 0} E_a \neq 0$  for both  $z$ - and  $\delta$ -models. Then formulae for error upper bounds of  $z$ - and  $\delta$ -models frequency response errors are valid:

$$\sup \left( \lim_{\lambda \rightarrow 0} E_z(\lambda) \right) = \frac{1.5A(\lambda) + 2}{|a(\lambda)| + 1.5\varepsilon \cdot \text{sgn}(a)} \varepsilon, \quad (4)$$

$$\sup \left( \lim_{\lambda \rightarrow 0} E_\delta(\lambda) \right) = \frac{2A(\lambda)}{1 + \varepsilon \cdot \text{sgn}(a)} \varepsilon. \quad (5)$$

Here and below  $E$  — error estimation,  $A = |H(\lambda)|$  — exact frequency response.

Proof of the lemma is based on the limit properties

$$\lim_{\lambda \rightarrow 0} z = 1,$$

$$\lim_{\lambda \rightarrow 0} \delta = 0,$$

and further considering the particular cases for the 1<sup>st</sup> and 2<sup>nd</sup> order sections. Other orders are impractical because first and

second order sections are common for linear systems implementation to avoid high round-off error in coefficient representation due to their wide range in high-order sections.

**Theorem 1.** Let complex frequency response for the frequency  $\lambda$  to be equal  $H_\delta(\lambda) = \beta/\alpha$  where  $\alpha, \beta$  — complex values for nominator and denominator of system transfer function  $H_\delta$ ; similarly  $H_z(\lambda) = b/a$ . Then when  $|\alpha| > K\varepsilon, \varepsilon < \varepsilon^*$ , where  $K, \varepsilon^*$  — known real numbers, the value

$$S_1 = \lim_{\lambda \rightarrow 0} \frac{\sup(E_z)}{\sup(E_\delta)}$$

does not depend on relative accuracy  $\varepsilon$ , i.e. on mantissa length of data type.

**Proof** is based on the fact that estimations  $\sup(E)$  contain  $\varepsilon$  as a coefficient, and division operation cancels it. Consider four probable cases.

1) Case  $|H(\lambda)| \neq 0, \lim_{\lambda \rightarrow 0} E_b \neq 0$

When condition  $|\alpha| > K\varepsilon$  on account of  $\varepsilon$  smallness is true, that  $|a| + 1.5\varepsilon \cdot \text{sgn}(a) \approx |a|$ , from where using (4)

$$\sup\left(\lim_{\lambda \rightarrow 0} E_z\right) = \frac{1.5A + 2}{|a|} \varepsilon,$$

and when  $K \gg \varepsilon^* > \varepsilon$  the formulation  $1 + \varepsilon \cdot \text{sgn}(a) \approx 1$  is correct, then

$$\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right) = 2A\varepsilon.$$

Therefore,

$$S_1 = \lim_{\lambda \rightarrow 0} \frac{\sup(E_z)}{\sup(E_\delta)} = \frac{\sup\left(\lim_{\lambda \rightarrow 0} E_z\right)}{\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right)} = \frac{\frac{1.5A + 2}{|a|} \varepsilon}{2A\varepsilon} = \frac{1.5}{|a|} + \frac{1}{|a|A}. \quad (6)$$

Note, that the equation (6) does not depend on  $\varepsilon$ .

2) Case  $A = 0, \lim_{\lambda \rightarrow 0} E_b \neq 0$  (case is representative e.g. for a notch filter).

Because of the inaccuracy of computer representation the real response values  $A_\delta \neq 0, A_z \neq 0$ .

For the z-system

$$\sup\left(\lim_{\lambda \rightarrow 0} E_z\right) = \frac{2}{|a|} \varepsilon,$$

if  $A_z \ll \varepsilon$ , that is quite likely from the case conditions.

For the  $\delta$ -system

$$\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right) = \frac{\varepsilon|\alpha_2\|\beta_2| + \varepsilon|\alpha_2\|\beta_2|}{\alpha_2(\alpha_2 + 1.5\varepsilon)} = 2A\varepsilon.$$

In the real system  $|\beta_2| \neq 0$ , though

$$\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right) = \lim_{\lambda \rightarrow 0} 2A_\delta\varepsilon = 0.$$

Therefore

$$S_1 = \lim_{\lambda \rightarrow 0} \frac{\sup(E_z)}{\sup(E_\delta)} = \frac{\sup\left(\lim_{\lambda \rightarrow 0} E_z\right)}{\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right)} = \frac{\frac{2}{|a|} \varepsilon}{2A_\delta\varepsilon} = \frac{1}{|a|A_\delta}$$

and does not depends on  $\varepsilon$ .

3) Case  $A \neq 0, \lim_{\lambda \rightarrow 0} E_b = 0$ , which is representative for low-pass and high-pass filters, where numerators converge to integer numbers. Then

$$\sup\left(\lim_{\lambda \rightarrow 0} E_z\right) = \frac{1.5A}{|a|} \varepsilon,$$

$$\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right) = \lim_{\lambda \rightarrow 0} A_\delta\varepsilon = 0,$$

and equation

$$S_1 = \lim_{\lambda \rightarrow 0} \frac{\sup(E_z)}{\sup(E_\delta)} = \frac{\sup\left(\lim_{\lambda \rightarrow 0} E_z\right)}{\sup\left(\lim_{\lambda \rightarrow 0} E_\delta\right)} = \frac{\frac{1.5A}{|a|} \varepsilon}{A\varepsilon} = \frac{1.5}{|a|},$$

does not depend on  $\varepsilon$ .

4) Case  $A = 0, \lim_{\lambda \rightarrow 0} E_b = 0$  corresponds the trivial condition  $H \equiv 0$ .

Conclusions of proven theorem have practical importance.

1) In most cases  $|\alpha| > 1.5\varepsilon$ , i.e.  $a_1$  differs from -2 by more than one least significant bit and  $a_2$  differs from 1 also by more than one bit. Consequently, relations of upper bounds of z- and  $\delta$ -system errors does not depend on  $\varepsilon$  or data type. With data precision growth accuracy increases for the both z- and  $\delta$ -model. Having  $\lambda$  and  $T$  specified there is no reason to increase mantissa length for z-model if  $\delta$ -model shows better performance on the same data type.

2). When extremely high frequencies or extremely low-precise data types are presented, i.e.  $|\alpha| < 1.5\varepsilon$ ,  $\delta$ -model preserves accuracy proportional to  $\varepsilon$

$$\lim_{\lambda \rightarrow 0} \sup(E_\delta) = 2A\varepsilon,$$

and z-model error is commensurable to the frequency response magnitude making it practically unusable:

$$\limsup_{\lambda \rightarrow 0}(E_z) = A + 1\frac{1}{3}.$$

In these cases  $\delta$ -operator makes possible to build discrete systems which cannot be implemented with usual techniques.

#### IV. THE PERFORMANCE OF DIFFERENT FILTERS WITH NORMALIZED FREQUENCY

Let us examine dependence of  $S$  and  $S_1$  on cutoff frequency  $\lambda$  for the 2<sup>nd</sup> order Butterworth lowpass filter implemented in 16-bit arithmetic. The feature of this filter type is exact definition of z-model numerator coefficients (they both converge to -1), so estimation

$$S_1 = \frac{1,5}{|a| + 1,5\epsilon}$$

is of interest. We obtained the formula through division of (4) by (5) and common factors cancellation.  $S_1$  estimation plot is shown on Fig. 4. Square sign notes the point  $|a| = 1,5\epsilon$ . One can see that when  $|a| > 1,5\epsilon$   $S_1$  decreases almost linearly as  $\lambda$  increases. When  $|a| < 1,5\epsilon$  the plot is horizontal because  $S_1$  does not depend on  $\lambda$ . The general shape of  $S_1$  approximately replicates  $S = E_z(\lambda)/E_\delta(\lambda)$  but when  $|a| > 1,5\epsilon$  estimation of  $S_1$  is overstating. Thus, estimation of  $S_1$  helped us to investigate behavior of criterion  $S$  but it could not be recommended for practical error evaluation.

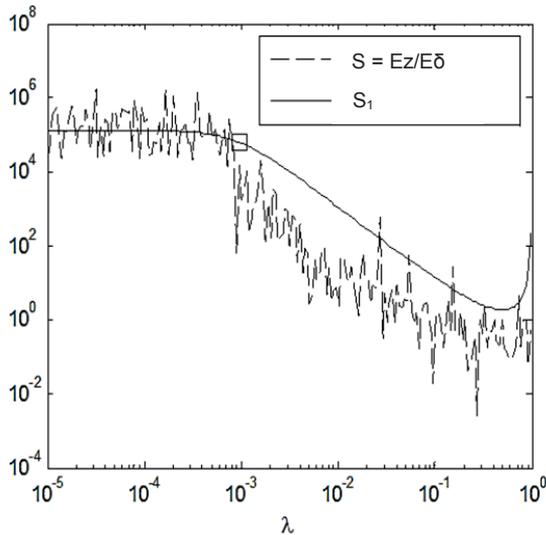


Fig. 4.  $S$  and  $S_1$  dependence to cutoff frequency  $\lambda$

Fig. 5 represents the numerical error estimations for both operators.

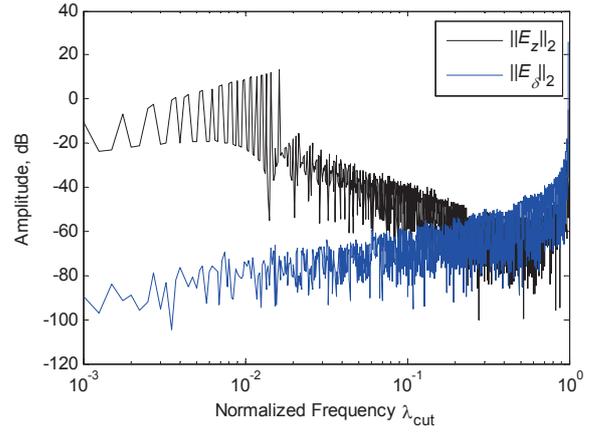


Fig. 5. The plot of  $\|E_z\|$  and  $\|E_\delta\|$

One can notice that in regions where  $\delta$ -operator is generally better z-operator shows better accuracy for some  $\lambda$  values. Local maxima of  $\delta$ -coefficients roundoff error and local minima of z-system error are located at these points. So variation of  $\lambda$  or, similarly, sampling rate or the specified cutoff frequency of the filter should be taken into account for development of the efficient discrete system synthesis algorithm.

Some experiments were performed to find  $\lambda^*$  values for different 2<sup>nd</sup> order filter types. Denote complex conjugate poles of the discrete system as  $r_{12} = \sigma \pm j\omega$  (obviously, it depends on  $\lambda$ ). It is found that value  $R(\lambda) = 1 - \|r_1(\lambda)\| = 1 - \|r_2(\lambda)\|$  can serve as another  $\delta$ -operator preference criterion. When  $R < 0.6$  filter implementation with  $\delta$ -operator shows better precision with probability of 95%. Fig. 6 represents this graphically.

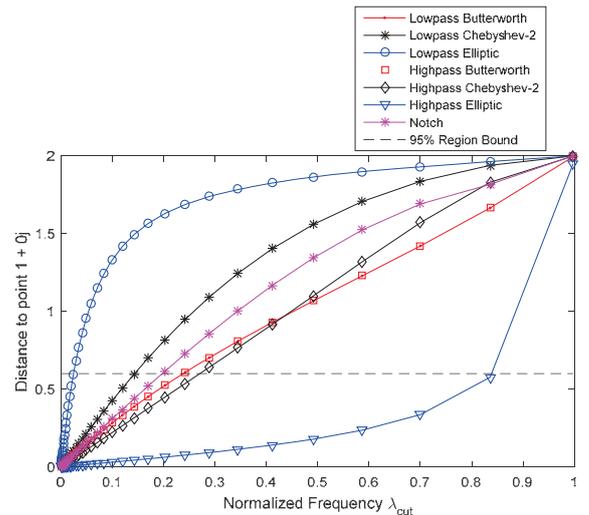


Fig. 6.  $R(\lambda)$  criterion for different filter types

However, this criterion is not applicable in general case of transfer functions with undefined cutoff frequency.

## V. ENHANCED ALGORITHM FOR DISCRETE OPERATOR SELECTION

To visualize general  $\delta$ -operator preference criterion we consider value

$$E = \frac{\|E_z\|}{\|E_\delta\|}$$

for systems obtained as a discrete transform of continuous systems with two complex conjugate poles  $r_{1,2} = \sigma \pm j\omega$  and transfer function

$$H(s) = \frac{1}{s^2 + a_1s + a_2}.$$

Using Tustin transform find  $H(z)$  and then  $H(\delta)$ . Coefficients are normalized in such a way that 1 remains a leading order variable held constant. Then a mantissa length was decreased to 8 bits (according to theorem 1, data type does not affect value  $S$ , so width was chosen randomly), frequency responses were calculated on the interval  $[0; \pi]$  and compared with double precision system frequency response.

As a result the surface  $S(\sigma T, \omega T)$  under the complex plane  $(\sigma T; \omega T)$  was built, where  $T$  denotes sampling time. In Fig. 7 the preference region bound line separates the level  $S=1$ . The points on the complex plane correspond to the values  $S > 1$  and determine the desired  $\delta$ -operator preference region. One observation makes  $\delta$ -operator preference criterion very simple: the most valuable part of the region is located inside the unit circle  $(\sigma T)^2 + (\omega T)^2 = 1$  with center in the origin. Statistical nature of the criterion admits this assumption.

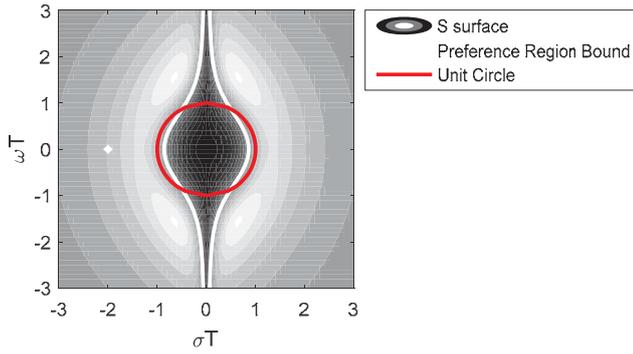


Fig. 7. Delta operator preference region on the complex continuous plane

Thus, an enhanced operator selection algorithm follows:

- 1) Decompose continuous system into second order sections.
- 2) Check the characteristic polynomial roots to be inside preference region  $S$ . If this condition is satisfied,  $\delta$ -operator must be chosen.

3) If sampling frequency  $\lambda$  variation is acceptable in certain neighborhood  $\lambda^*$ , find the local minimum and maximum for criterion  $S$ . If

$$\frac{1}{S_{\min}} < S_{\max},$$

$\delta$ -operator must be chosen, else z-operator. If precise value of  $S$  is impossible to find (estimation of  $S$  is calculated), then  $\delta$ -operator must be chosen using condition  $S > 1$ .

Let us consider 2<sup>nd</sup> order low-pass elliptic filter with the following specs: passband ripple  $10^{-4}$  dB, stopband attenuation 40 dB, normalized passband edge  $\lambda_{cut} = 0.03$  (case 1) and  $\lambda_{cut} = 0.002$  (case 2) with  $S(\sigma T, \omega T)$ -criterion. Remind these specs from section 1 and formulae (1) and (2). Continuous filter equation using  $f_{samp} = 1$  kHz with  $\lambda_{cut} = 0.03$  is:

$$H_1(s) = \frac{0.01s^2 + 3.7022 \cdot 10^6}{s^2 + 2.709 \cdot 10^3 s + 3.7023 \cdot 10^6}.$$

The polynomial roots of filter denominator are  $r_{1,2} = -1350.5 \pm 1370.6i$ . The point  $(-1.35, -1.37)$  is not in the unit circle and z-operator is preferable. When  $\lambda_{cut} = 0.002$

$$H_2(s) = \frac{0.01s^2 + 1.645 \cdot 10^4}{s^2 + 1.806 \cdot 10^2 s + 1.6454 \cdot 10^4},$$

denominator polynomial roots are  $r_{1,2} = -90.03 \pm 91.372i$ , and  $(-0.09, -0.091)$  indicates that system will have better performance with  $\delta$ -operator. These results completely conform to data from section 1.

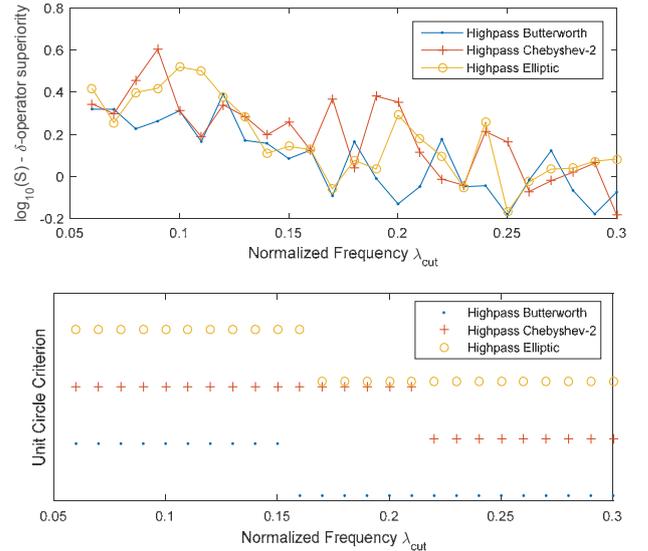


Fig. 8. Logarithmic error relation  $\log_{10} S$  and unit circle criterion response

For the study extension filters from section 4 were implemented in floating-point arithmetic using z- and  $\delta$ -operator. Fig.8 illustrates some of the highpass filters tests. One can see the correspondence of two plots (a) and (b): criterion steps off near the points where  $\log_{10} S$  lines cross zero line the first time. This matches frequencies where  $\delta$ -systems start to lose their advantage over z-systems.

## VI. CONCLUSION

We investigated the influence of normalized cutoff frequency  $\lambda$ , filter type and machine arithmetic precision on discrete dynamical models accuracy when implementing models with z- and  $\delta$ -operators. It was found that there always exists a specific cutoff frequency  $\lambda < \lambda^*$  where  $\delta$ -operator model becomes more precise than z-model regardless to data type. Though this enhancement may not be very appreciable in absolute values, e.g. the magnitude error of frequency response  $10^{-5}$  for z-model and  $10^{-8}$  for  $\delta$ -model, when low-precision data representation is used such thousandfold superiority significantly increases accuracy. We proposed a simple criterion to answer the question which discrete operator is preferable for a given continuous model and discretization time. This criterion was examined experimentally on a set of IIR filters. The research of obtained z/ $\delta$ -models showed that in a most cases proposed choice criterion is valid. Future work will be dedicated to the development of the best fixed-point scaling technique for  $\delta$ -models.

## ACKNOWLEDGEMENTS

The reported study was supported by RFBR, research project No. 17-57-52017.

## REFERENCES

- [1] G. Govindu, L. Zhuo, S. Choi, and V. Prasanna, "Analysis of High-performance Floating-point Arithmetic on FPGAs", *18th International Parallel and Distributed Processing Symposium (IPDPS)*, 2004, pp. 1–8.
- [2] A. I. Karimov and T. I. Karimov, "Dynamic system simulation using delta operator in LabVIEW environment", *Proceedings of the 2015 International Siberian Conference on Control and Communications (SIBCON)*, Omsk, Russia, 2015.
- [3] Y. Liu, Y. Wang, Z. Gao, and C. Dou, "Optimal sliding mode control for delta operator system", *IEEE 32nd Chinese Control Conference*, 2013, pp. 3152 – 3157.
- [4] B. L. Eidson, J. Y. Hung, and R. Mark Nelms, "An experimental evaluation of the PID controller represented by the delta operator", *Proceedings of IEEE Southeastcon*, Mar. 2012, pp. 1–6.
- [5] S. Longo, E. C. Kerrigan, and G. A. Constantinides, "Constrained LQR for low-precision data representation", *Automatica*, vol. 50, no. 1, 2014, pp. 162–168.
- [6] G. H. Yang, *Linear systems : non-fragile control and filtering*. CRC Press, 2013.
- [7] C. A. Rabbath and N. Léchevin, *Discrete-Time Control System Design with Applications*. New York: Springer, 2014.
- [8] R. M. Goodall and B. J. Donoghue, "Very high sample rate digital filters using the  $\delta$  operator", *IEEE Proceedings of Circuits, Devices and Systems*, vol. 140, no. 3, 1993, pp. 199–208.
- [9] T. I. Karimov, D. N. Butusov, and A. I. Karimov, "The criterion of delta-operator application in state space digital systems design", *Modern problems of science and education*, vol. 12, no. 9, 2014, pp. 1889–1893.
- [10] R. H. Middleton and G. C. Goodwin, "Improved Finite Word Length Characteristics in Digital Control Using Delta Operators", *IEEE Transactions on Automatic Control*, vol. 31, Nov. 1986, pp. 1015–1021.
- [11] M. J. Newman and D. G. Holmes, "Delta operator digital filters for high performance inverter applications", *IEEE Transactions on Power Electronics*, vol. 18, no. 1, Jan. 2003, pp. 447–454.
- [12] J. Kauraniemi, T. I. Laakso, and I. Hartimo, "Delta Operator Realizations of Direct-Form IIR Filters", *IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 45, no. 1, 1998, pp. 41–52.