Московский государственный технический университет им.Н.Э.Баумана
Калужский филиал
Приборостроительный факультет
Кафедра П2-КФ

Отчет

по лабораторной работе № 2

«Исследование инструментальной погрешности реализации функций»

по курсу

«Проектирование СЦВМ»

Выполнил: студент Комиссаров А.В., гр.ЭВМ-112

Принял : доцент Максимов А.В.

Калуга, 1999 г.

<u>Задание (вариант № 6)</u>

На вход ЦВМ поступает информация по двум каналам:

$$x = f(t) = 0{,}25 \cdot t^2 + 5;$$
$$y = \varphi(t) = 5 \cdot \sin(t^2);$$
$$t = 0..20 \text{ сек.}$$

Входная информация, поступающая по каждому из каналов, обрабатывается по своему алгоритму:

$$u(x) = 2 \cdot x^4 + 0{,}5 \cdot x^3 + \cdot x^2 + x + 6;$$
$$w(y) = 0{,}5 \cdot \exp(-y) \cdot \sin(y).$$

<u>Дополнительные указания :</u>

- число разрядов сетки $N = 16$;
- необходимо взять первые три ненулевых члена разложения функции в ряд Маклорена.

<u>Требуется :</u>

- разработать математический, машинный алгоритмы реализации функций обработки и смоделировать вычисление функций u и w (чтобы избежать появления трансформированной ошибки, рекомендуется подавать на вход только значения, полностью входящие в разрядную сетку);
- определить практические значения инструментальной погрешности для каждого из каналов, а также дисперсии и среднеквадратичного отклонения инструментальной погрешности;
- определить теоретические значения инструментальной погрешности, а также дисперсии и среднеквадратичного отклонения инструментальной погрешности;
- сопоставить теоретические значения с практическими и сделать выводы.

Ход лабораторной работы

1. Канал *u(x)*

1.1. Математический алгоритм

Более оптимальным (по сравнению с исходной формой) является следующая реализация функции $u(x)$:

$$u(x) = x \cdot \left(1 + x \cdot \left(1 + x \cdot \left(2^{-1} + 2 \cdot x\right)\right)\right) + 6.$$

Получаем следующий **математический алгоритм** :

$$u_1 = 2 \cdot x;\quad u_2 = 2^{-1} + u_1;\; u_3 = x \cdot u_2;\; u_4 = 1 + u_3;$$

$$u_5 = x \cdot u_4;\; u_6 = 1 + u_5;\; u_7 = x \cdot u_6;\; u_8 = 6 + u_7\ .$$

1.2. Машинный алгоритм и масштабные соотношения. Теоретический расчет инструментальной погрешности

Представим **машинный алгоритм**, а также **масштабные соотношения** для каждой из операций в виде таблицы (табл. 1). Каждая из операций реализуется с учетом всех возможностей по повышению точности. В крайней правой колонке таблицы приведены теоретические соотношения и определены теоретические значения максимальных погрешностей, а также дисперсий для каждой из операций. Теоретические расчеты были проведены с помощью системы MathCad 7.0 Professional Edition.

Сделаем замечание. Для правильного моделирования работы контура САУ с ЦВМ в цепи обратной связи на языке высокого уровня (например, на языке Delphi Pascal) необходимо осуществить перевод дробных масштабов аргумента и результата в целочисленные. Для этого необходимо соответствующие дробные двоичные масштабы умножить на величину $2^n = 2^{16} = 65536$ (так как функция $u(x)$ положительна на рассматриваемом интервале x = 0…105, знаковый разряд можно не вводить). Таким образом, получаем :

$$M_X = 2^{-7} \cdot 2^{16} = 2^9 = 512;$$

$$M_U = 2^{-28} \cdot 2^{16} = 2^{-12} = 1/4096.$$

Используемые в алгоритме машинные значения констант: $2^{-15} = 0002h;\; 2^{-9} = 0080h.$

Теоретические зависимости получены из формул, известных из теории вероятностей и математической статистики, а также с учетом того, что:

- инструментальная погрешность усечения переменной равна сумме отбрасываемых разрядов и распределена по *равномерному* закону;

- дисперсия постоянной ошибки (получаемой в результате представления констант) равна нулю, даже если сама постоянная ошибка отлична от нуля;

- инструментальная ошибка, полученная на каждой из операций, трансформируется тем или иным образом на последующих операциях.

Таблица 1. Машинные алгоритмы и масштабные соотношения для каждой операции

Операция	Машинный алгоритм	Масштабные соотношения и определение макс. знач.	Расчет погрешности и ее дисперсии (в истинных величинах)
$u_1 = 2 \cdot x$	$\bar{u}_1 := \bar{x}$	$\mathcal{M}_X = 2^{-7}$ $\mathcal{M}_{u_1} = 2^{-7} \cdot 2^{-1} = 2^{-8}$ $\lvert u_1 \rvert_{\max} = 210$	$\Delta u_1 = \Delta x = 0;\ D(\Delta u_1) = 0$
$u_2 = 2^{-1} + u_1$	$\bar{u}_2 := \left(2^{-1} \cdot 2^{-8}\right) + \bar{u}_1$	$\mathcal{M}_{0,5} = 1$ $\mathcal{M}_{u1} = 2^{-8};\ \lvert u_2 \rvert_{\max} = 210{,}5$ $\mathcal{M}_{u2} = \min\{2^{-8}; 2^{0}; 2^{-8}\} = 2^{-8}$ Коэффициенты выравнивания $K_{0,5} = \frac{\mathcal{M}_{u2}}{\mathcal{M}_{0,5}} = \frac{2^{-8}}{2^{0}} = 2^{-8}$ $K_{u1} = 1$	$\Delta u_2 = \Delta\left(2^{-1}\right) + \Delta u_1 = 0;$ $D(\Delta u_2) = 0$
$u_3 = x \cdot u_2$	$\bar{u}_3 = \bar{x} \cdot \bar{u}_2$	$\mathcal{M}_X = 2^{-7}; \mathcal{M}_{u2} = 2^{-8};$ $\lvert u_3 \rvert_{\max} = 105 \cdot 210{,}5 = 22102{,}5$ $\mathcal{M}_{u3}(22102{,}5) = 2^{-15}$ $k = \log_2\left(\frac{2^{-15}}{2^{-7} \cdot 2^{-8}}\right) = 0^{*}$	$\Delta u_{3\max} = \sum_{i=17}^{32} 2^{-i} / 2^{-15} = S_1;$ ** $D(\Delta u_3) = \frac{1}{3} \cdot \Delta^2 u_{3\max} = \frac{1}{3} \cdot S_1^2$
$u_4 = 1 + u_3$	$\bar{u}_4 := \left(2^{-1} \cdot 2^{-14}\right) + \bar{u}_3$	$\mathcal{M}_1 = 2^{-1};\ \mathcal{M}_{u3} = 2^{-15};$ $\lvert u_4 \rvert_{\max} = 22103{,}5$ $\mathcal{M}_{u4} = \min\{2^{-1}; 2^{-15}; 2^{-15}\} = 2^{-15}$ Коэффициенты выравнивания $K_1 = \frac{\mathcal{M}_{u4}}{\mathcal{M}_1} = \frac{2^{-15}}{2^{-1}} = 2^{-14}$ $K_{u3} = 1$	$\Delta u_{4\max} = \Delta u_{3\max} = S_1;$ $D(\Delta u_4) = \frac{1}{3} \cdot \Delta^2 u_{4\max} = \frac{1}{3} \cdot S_1^2$ Заметим, что представление константы 2^{-15} не дает ошибки.
$u_5 = x \cdot u_4$	$\bar{u}_5 = \bar{x} \cdot \bar{u}_4$	$\mathcal{M}_X = 2^{-7}; \mathcal{M}_{u4} = 2^{-15};$ $\lvert u_5 \rvert_{\max} = 2320867{,}5$ $\mathcal{M}_{u5}(\max) = 2^{-22}$ $k = \log_2\left(\frac{2^{-22}}{2^{-7} \cdot 2^{-15}}\right) = 0$	$\Delta u_{5\max} = \Delta u_{4\max} x_{\max} + \Delta u_{5\ ус.\max};$ $\Delta u_{5\ ус.\max} = \sum_{i=17}^{32} 2^{-i} / 2^{-22} = S_2$ ***; $D(\Delta u_5) = x_{\max}^2 \cdot D(\Delta u_4) + \frac{1}{3} S_2^2$

* Так как знаковый бит отсутствует, изменились и формулы для расчета масштабов. Так, теперь не нужно умножать масштаб результата при умножении на 2^{-1}. k – число сдвигов результата влево.

** $S_1 = 0{,}4999923\ldots$ Значения подобных сумм могут быть найдены через формулы для суммы конечного числа членов геометрической прогрессии со знаменателем прогрессии $q = 2^{-1}$

*** $S_2 = 63{,}999023$

Таблица 1. Машинные алгоритмы и масштабные соотношения для каждой операции (продолжение)

Операция	Машинный алгоритм	Масштабные соотношения и определение макс. знач.	Расчет погрешности и ее дисперсии (в истинных величинах)
$u_6 = 1 + u_5$	$\bar{u}_6 := \left(2^{-1}\cdot 2^{-21}\right) + \bar{u}_5$ Константа 2^{-22} не представима в 16-разрядной сетке, поэтому в данном случае $\bar{u}_6 := \bar{u}_5$	$\mathcal{M}_1 = 2^{-1}$; $\mathcal{M}_{u5} = 2^{-22}$; $\lvert u_6 \rvert_{\max} = 2320868{,}5 < 2^{22}$ $\mathcal{M}_{u6} = \min\{2^{-1}; 2^{-22}; 2^{-22}\} = 2^{-22}$ Коэффициенты выравнивания $K_1 = \frac{\mathcal{M}_{u6}}{\mathcal{M}_1} = \frac{2^{-22}}{2^{-1}} = 2^{-21}$ $K_{u5} = 1$	$\Delta u_{6\,\max} = 2^{-21} + \Delta u_{5\,\max}$; $\Delta u_{6\,\max} = 2^{-21} + 105\cdot S_1 + S_2$; $D(\Delta u_6) = 0 + D(\Delta u_5)$[*] $D(\Delta u_6) = 11025\cdot\frac{1}{3}\cdot S_1^2 + \frac{1}{3}\cdot S_2^2$
$u_7 = x \cdot u_6$	$\bar{u}_7' := \bar{x}\cdot\bar{u}_6$ $u_7 := \bar{u}_7' \cdot 2^1$	$\mathcal{M}_x = 2^{-7}$; $\mathcal{M}_{u6} = 2^{-22}$; $\lvert u_7 \rvert_{\max} = 2{,}437\cdot 10^8$ $\mathcal{M}_{u7}(\max) = 2^{-28}$ $k = \log_2\left(\frac{2^{-28}}{2^{-7}\cdot 2^{-22}}\right) = 1$	$\Delta u_{7\,\max} = \Delta u_{6\,\max} x_{\max} + \Delta u_{7\,\text{ус.}\max}$; $\Delta u_{7\,\text{ус.}\max} = \sum_{i=17}^{31} 2^{-i}/2^{-28} = S_3$[***]; $D(\Delta u_7) = x_{\max}^2 \cdot D(\Delta u_6) + \frac{1}{3} S_3^2$
$u_8 = 6 + u_7$	$\bar{u}_8 := \left(2^{-26} + 2^{-27}\right) + \bar{u}_7$ Константа ($2^{-26} + 2^{-27}$) не представима в 16-разрядной сетке, поэтому в данном случае $\bar{u}_8 := \bar{u}_7$	$\mathcal{M}_6 = 2^{-3}$; $\mathcal{M}_{u7} = 2^{-28}$; $\lvert u_8 \rvert_{\max} = \left(2{,}437\cdot 10^8\right) + 6 < 2^{28}$ $\mathcal{M}_{u8} = \min\{2^{-3}; 2^{-28}; 2^{-28}\} = 2^{-28}$ Коэффициенты выравнивания $K_6 = \frac{\mathcal{M}_{u8}}{\mathcal{M}_6} = \frac{2^{-28}}{2^{-3}} = 2^{-25}$ $K_{u7} = 1$	$\Delta u_{8\,\max} = \Delta u_{7\,\max} + \Delta_{\text{ус.}}\left(\bar{6}K_6\right)$; $\Delta_{\text{ус.}}\left(\bar{6}K_6\right) = \left(2^1 + 2^2\right)\cdot 2^{-3}\cdot 2^{-25}/2^{-28}$; $\Delta_{\text{ус.}}\left(\bar{6}K_6\right) = 2^1 + 2^2 = 6$; $D(\Delta u_8) = D(\Delta u_7)$.

1.3. Экспериментальный расчет математического ожидания и дисперсии ошибки

Заметим, что при практической реализации алгоритма значения будем подавать с физическим шагом, кратным промасштабированному минимальному из представимых в разрядной сетке значению аргумента (в данном случае это $(2^{-16}/2^{-7})\cdot 64 = 0{,}125$). Здесь 64 – коэффициент, увеличивающий значение шага и скорость моделирования. Это один из тех случаев, когда исходные аргументы представляются точно в разрядной сетке.

В качестве эталонной функции возьмем ее реализацию в системе Delphi, где и производилось моделирование (тип данных extended, вещественный формат, 10 байт, 19..20 значащих цифр мантиссы).

Ошибка между *i*-м значением реализации функции на языке ассемблера и значением эталонной функции может быть вычислена следующим образом :

$$E_i = y_i - f_i ,$$

где y_i - текущее значение реализуемой функции, а f_i - текущее значение эталонной функции (первые четыре члена ряда).

[*] Дисперсия постоянной равна нулю
[***] $S_3 = 4095{,}875$

Математическое ожидание ошибки реализации можно подсчитать по формуле :

$$M[E] = \frac{\sum_i E_i}{n},$$

где n – количество точек, в которых вычисляется значение функции на заданном интервале.

Дисперсию (квадрат среднеквадратического отклонения) ошибки реализации можно подсчитать по формуле :

$$D[E] = M[E^2] - M^2[E].$$

Среднеквадратичное отклонение есть квадратный корень от дисперсии.

Ниже приводятся результаты теоретических расчетов, а также экспериментальные значения инструментальной погрешности и ее числовых характеристик.

Таблица 2. Исследование инструментальной погрешности ΔU(x)

Величина	Экспериментальное значение	Теоретическое (расчетное) значение
Максимальная погрешность	14678	16334
Средняя погрешность	4290	Не расчитывалось
Дисперсия погрешности	$6747 \cdot 10^3$	$30770 \cdot 10^3$
Среднекв. откл. погрешности	2598	5547

1.4. Фрагмент программы, моделирующий вычислительный блок U(x) СЦВМ (ассемблерная вставка (CPU ix86 ASM) в Delphi Pascal - модуль)

```
function AsmCalcU(X : word) : word;
const
    const_2m9 = $0080;
    const_2m15 = $0002;
var
  U : word;
begin
   asm
     pusha
     pushf
     mov   ax, X
     mov   bx, ax
     add   ax, const_2m9   //z2 in AX
     mul   bx              //z3 in DX:AX
     add   dx, const_2m15  //z4 in DX
     mov   ax, dx          //z4 in AX
     mul   bx              //z5 in DX:AX
     //z6 = 2m22 + z5... N=16 <=> z6 = z5 in DX:AX
     mov   ax, dx
     mul   bx
     shl   dx, 1
     shl   ax, 1
     adc   dx, 0
     //z7 in DX:AX; u=z8=z7
     mov   U, dx
     popf
     popa
   end;
   Result := U;
end;
```

2. Канал *w(y)*

2.1. Математический алгоритм

Разложения исходных составляющих функции $w(y)$, $w_1(y) = \exp(-y)$ и $w_2(y) = \sin(y)$, в ряды Маклорена, в которых взяты только первые три члена, соответственно имеют вид :

$$w_1(y) \approx 1 - y + \frac{y^2}{2} = 1 - y \cdot \left(1 - \frac{y}{2}\right);$$

$$w_2(y) \approx y - \frac{y^3}{6} + \frac{y^5}{120} = y \cdot \left(1 - \frac{y^2}{6} \cdot \left(1 - \frac{y^2}{20}\right)\right).$$

Разложение функции sin(y) в ряд даст наименьшую ошибку сходимости на интервале $y \in \left[-\frac{\pi}{4}; \frac{\pi}{4}\right]$. В данной лабораторной работе осуществляется приведение аргумента к интервалу $y \in \left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$, так как в этом случае алгоритм приведения и реализации функции получается значительно более простым (в противном случае пришлось бы дополнительно реализовывать функцию cos(y) и различные сочетания суммы взвешенных sin и cos).

Итак, $w(y) = 0{,}5\ w_1(y)\ w_2(y_1)$, где приведенный аргумент y_1 равен (для обоснования следует рассмотреть изменение значения аргумента, находящегося в каждом из четырех квадрантов и принимающего как положительные, так и отрицательные значения в пределах от 0 до ±5 таким образом, что $\sin(y_1) = \sin(y)$):

$$(*)\ y_1 = \begin{cases} y,\ если \quad |y| \le \frac{\pi}{2}; \\ \pi - y,\ если\ \frac{\pi}{2} < y \le \frac{3\pi}{2}; \\ y - 2\pi, если\ \frac{3\pi}{2} < y \le 2\pi; \\ -y - \pi, если\ \frac{-3\pi}{2} \le y < -\frac{\pi}{2}; \\ y + 2\pi,\ если\ -2\pi \le y < \frac{-3\pi}{2}. \end{cases}$$

2.2. Машинный алгоритм и масштабные соотношения. Теоретический и экспериментальный расчет инструментальной погрешности

Представим **машинный алгоритм**, а также **масштабные соотношения** для каждой из операций в виде таблиц (табл. 3.1 – 3.4).

В таблице всюду использована дробная арифметика. Заметим однако, что при моделировании в системе Delphi фактически приходится иметь дело с целыми числами (типа smallint, -32768…32767, знаковые 16-битные целые числа), поэтому соответствующие дробные масштабы аргументов и результата переводятся в целочисленные путем умножения на $2^n = 2^{15}$.

Таблица 3.1. Машинные алгоритмы и масштабные соотношения для каждой операции. Вычисление $w_1(y)$

Операция	**Машинный алгоритм**	**Масштабные соотношения и определение макс. знач.**	**Расчет погрешности и ее дисперсии (в истинных величинах)**
$z_1 = 2^{-1} \cdot y$	$\bar{z}_1 := \bar{y}$	$\mathcal{M}_y = 2^{-3}$; $\lvert y \rvert_{\max} = 5$; $\mathcal{M}_{z1} = 2^{-3} \cdot 2^1 = 2^{-2}$ $\lvert z_1 \rvert_{\max} = 2{,}5$	$\Delta z_1 = \Delta y = 0;\ D(\Delta z_1) = 0$
$z_2 = 1 - z_1$	$\bar{z}_2 := \left(2^{-1} \cdot 2^{-1}\right) - \bar{z}_1$	$\mathcal{M}_1 = 2^{-1}$; $\mathcal{M}_{z1} = 2^{-2}$; $\lvert z_2 \rvert_{\max} = 3{,}5 < 2^2$; $\mathcal{M}_{z2} = \min\left\{2^{-1}; 2^{-2}; 2^{-2}\right\} = 2^{-2}$ Коэффициенты выравнивания: $K_{z1} = \dfrac{\mathcal{M}_{z2}}{\mathcal{M}_{z1}} = \dfrac{2^{-2}}{2^{-2}} = 1;$ $K_1 = 2^{-1}$	$\Delta z_2 = \Delta\left(\bar{1} \cdot 2^{-1}\right) + \Delta z_1 = 0;$ $D(\Delta z_2) = 0$
$z_3 = y \cdot z_2$	$\bar{z}_3 = 2^1\left(\bar{y} \cdot \bar{z}_2\right)$	$\mathcal{M}_y = 2^{-3}$; $\mathcal{M}_{z2} = 2^{-2}$; $\lvert z_3 \rvert_{\max} = \lvert -5 \cdot 3{,}5 \rvert = 17{,}5$ $\mathcal{M}_{z3}(\max) = 2^{-5}$; $k = \log_2\left(\dfrac{2^{-5}}{2^{-3} \cdot 2^{-2} \cdot 2^{-1}}\right) = 1$	$\Delta z_{3\max} = \sum_{i=16}^{30} 2^{-i} / 2^{-5} = \Sigma_1;$ $D(\Delta z_3) = \dfrac{1}{3} \cdot \Delta^2 z_{3\max} = \dfrac{1}{3} \cdot \Sigma_1^2$
$w_1 = z_4 = 1 - z_3$	$\bar{z}_4 := \left(2^{-1} \cdot 2^{-4}\right) - \bar{z}_3$	$\mathcal{M}_1 = 2^{-1}$; $\mathcal{M}_{z3} = 2^{-5}$; $\lvert z_4 \rvert_{\max} = 18{,}5$; $\mathcal{M}_{z4} = \min\left\{2^{-1}; 2^{-5}; 2^{-5}\right\} = 2^{-5}$; Коэффициенты выравнивания:; $K_{z3} = 1$.	$\Delta z_{4\max} = \Delta z_{3\max} = S_1;$ $D(\Delta z_4) = \dfrac{1}{3} \cdot \Delta^2 z_{4\max} = \dfrac{1}{3} \cdot S_1^2$ Заметим, что представление константы 2^{-5} не дает ошибки.

Таблица 3.2. Машинные алгоритмы и масштабные соотношения для каждой операции. Вычисление аргумента для $w_2(y)$ – см. систему уравнений (*)

Операция	Машинный алгоритм	Масштабные соотношения и определение макс. знач.
$y_1 = y$	$\bar{y}_1 := (2^2)\ \bar{y}$	$\mathcal{M}_y = 2^{-3}$; $\lvert y_1 \rvert_{\max} = \dfrac{\pi}{2} < 2^1$ Данный масштаб неоптимален с точки зрения максимума результата, поэтому вводится коррекция: $\mathcal{M}_{y1} = 2^{-1}$
$y_1 = \pi - y$	$\bar{y}_1 := (2^2)\left((\bar{\pi} \cdot 2^{-1}) - \bar{y}\right)$	$\mathcal{M}_{y1}^{\sim} = \min\{2^{-2}; 2^{-3}; 2^{-1}\} = 2^{-3}$ Данный масштаб неоптимален с точки зрения максимума результата, поэтому вводится коррекция: $\mathcal{M}_{y1} = 2^2 \cdot \mathcal{M}_{y1}^{\sim} = 2^{-1}$
$y_1 = y - 2\pi$	$\bar{y}_1 := (2^2)\left(\bar{y} - (\overline{2\pi})\right)$	$\mathcal{M}_{y1}^{\sim} = \min\{2^{-3}; 2^{-3}; 2^{-1}\} = 2^{-3}$ Данный масштаб неоптимален с точки зрения максимума результата, поэтому вводится коррекция: $\mathcal{M}_{y1} = 2^2 \cdot \mathcal{M}_{y1}^{\sim} = 2^{-1}$
$y_1 = -y - \pi$	$\bar{y}_1 := (2^2)\left(-\bar{y} - \bar{\pi} \cdot 2^{-1}\right)$	$\mathcal{M}_{y1}^{\sim} = \min\{2^{-3}; 2^{-2}; 2^{-1}\} = 2^{-3}$ Данный масштаб неоптимален с точки зрения максимума результата, поэтому вводится коррекция: $\mathcal{M}_{y1} = 2^2 \cdot \mathcal{M}_{y1}^{\sim} = 2^{-1}$
$y_1 = y + 2\pi$	$\bar{y}_1 := (2^2)\left(\bar{y} + \overline{2\pi}\right)$	$\mathcal{M}_{y1}^{\sim} = \min\{2^{-3}; 2^{-3}; 2^{-1}\} = 2^{-3}$ Данный масштаб неоптимален с точки зрения максимума результата, поэтому вводится коррекция: $\mathcal{M}_{y1} = 2^2 \cdot \mathcal{M}_{y1}^{\sim} = 2^{-1}$

В таблице 3.2 приведены машинные алгоритмы и масштабные соотношения для операций приведения аргумента. Так как отдельные случаи приведения соответствуют разным диапазонам изменения аргумента (т.е. ошибка, возникающая в результате приведения, появляется не на всем диапазоне), а теоретический расчет производится исходя из наиболее критичных ситуаций (считается, что в отбрасываемых и усекаемых разрядах переменных располагаются одни единицы, а коэффициенты трансформации равны максимальным значениям соответствующих частных производных), будем в аналитических расчетах условно полагать, что ошибка приведения* равна нулю.

---

* Эта ошибка складывается в общем случае из ошибки представления иррационального числа π (или 2π) в разрядной сетке, а также (в некоторых случаях) из ошибки сдвига на один разряд вправо представленного числа π в результате выравнивания масштабов при сложении (вычитании).

Таблица 3.3. Машинные алгоритмы и масштабные соотношения. Вычисление $w_2(y)$

Операция	Машинный алгоритм	Масштабные соотношения и определение макс. знач.	Расчет погрешности и ее дисперсии (в истинных величинах)
$s_1 = y_1 \cdot y_1$	$\bar{s}_1 = (\bar{y}_1 \cdot \bar{y}_1) \cdot 2^1$	$\mathcal{M}_{y1} = 2^{-1}$; $\lvert s_1 \rvert_{\max} = 2{,}467 < 2^2$; $\mathcal{M}_{s1}(\max) = 2^{-2}$; $k = \log_2\left(\frac{2^{-2}}{2^{-1} \cdot 2^{-1} \cdot 2^{-1}}\right) = 1$	$\Delta s_1 = 2\Delta y_1 \lvert y_1 \rvert_{\max} + \sum_{i=16}^{30} 2^{-i} / 2^{-2}$ $\Delta s_1 = 2\Delta y_1 \lvert y_1 \rvert_{\max} + \Sigma_2$ $D(\Delta s_1) = 4D(\Delta y_1) \lvert y_1 \rvert^2_{\max} + \frac{1}{3}\Sigma_2^2$
$s_2 = 0{,}05\ s_1$	$\bar{s}_2 = \left(\overline{\left(\frac{1}{20}\right)} \cdot \bar{y}_1\right) \cdot 2^2$	$\mathcal{M}_{s1} = 2^{-2}$; $\mathcal{M}_{s2}(\max) = 2^3$; $k = \log_2\left(\frac{2^{3}}{2^{4} \cdot 2^{-2} \cdot 2^{-1}}\right) = 1$	$\Delta s_2 = \frac{1}{20}\Delta s_1 + 2^{-16} \cdot \lvert s_1 \rvert_{\max} + \Sigma_3$ $\Sigma_3 = \sum_{i=16}^{29} 2^{-i} / 2^{3}$ $D(\Delta s_2) = \frac{1}{400} D(\Delta s_1) + \frac{1}{3}\Sigma_3^2$
$s_3 = 1 - s_2$	$\bar{s}_3 := 2^{-1} - (\bar{s}_2 \cdot 2^{-4})$	$\mathcal{M}_{s2} = 2^3$; $\mathcal{M}_1 = 2^{-1}$; $\lvert s_3 \rvert_{\max} = 1 < 2^1$; $\mathcal{M}_{s3} = \min\{2^{-1}; 2^3; 2^{-1}\} = 2^{-1}$	$\Delta s_3 = 2^{-4}\Delta s_2 + \sum_{i=16}^{19} 2^{-i} / 2^{3-4}$ (ошибка представления константы 2^{-1} равна нулю) $\Sigma_4 = \sum_{i=16}^{19} 2^{-i} / 2^{3-4}$; $D(\Delta s_3) = 2^{-8} D(\Delta s_2) + \frac{1}{3}\Sigma_4^2$
$s_4 = s_1\ s_3$	$\bar{s}_4 := (\bar{s}_1 \cdot \bar{s}_3) \cdot 2^2$	$\mathcal{M}_{s1} = 2^{-2}$; $\mathcal{M}_{s3} = 2^{-1}$ $\lvert s_4 \rvert_{\max} = 2{,}163 < 2^2$; $\mathcal{M}_{s4}(\max) = 2^{-2}$	$\Delta s_4 = \lvert s_1 \rvert_{\max} \Delta s_3 + \lvert s_3 \rvert_{\max} \Delta s_1 + \Sigma_5$ $\Sigma_5 = \sum_{i=16}^{29} 2^{-i} / 2^{-2}$ $D(\Delta s_4) = \lvert s_1 \rvert^2_{\max} \Delta s_3 + \lvert s_3 \rvert^2_{\max} \Delta s_1 + \frac{1}{3}\Sigma_5^2$
$s_5 = \frac{1}{6} \cdot s_4$;	$\bar{s}_5 := \left(\overline{\left(\frac{1}{6}\right)} \cdot \bar{s}_4\right) \cdot 2^2$	$\mathcal{M}_{s4} = 2^{-2}$; $\lvert s_5 \rvert_{\max} = 0{,}36 < 2^{-1}$ $\mathcal{M}_{\frac{1}{6}} = 2^2$; $\mathcal{M}_{s5}(\max) = 2^1$	$\Delta s_5 = \frac{1}{6}\Delta s_4 + \lvert s_4 \rvert_{\max} 2^{-16} + \Sigma_6$ $\Sigma_6 = \sum_{i=16}^{29} 2^{-i} / 2^{1}$; $D(\Delta s_5) = \frac{1}{36} D(\Delta s_4) + \frac{1}{3}\Sigma_6^2$
$s_6 = 1 - s_5$	$\bar{s}_6 := 2^{-1} - (\bar{s}_5 \cdot 2^{-2})$	$\mathcal{M}_{s5} = 2^1$; $\mathcal{M}_1 = 2^{-1}$; $\lvert s_6 \rvert_{\max} = 1 < 2^1$; $\mathcal{M}_{s6} = \min\{2^{-1}; 2^1; 2^{-1}\} = 2^{-1}$	$\Delta s_6 = 2^{-2}\Delta s_5 + \sum_{i=16}^{17} 2^{-i} / 2^{1-2}$ (ошибка представления константы 2^{-1} равна нулю) $\Sigma_7 = \sum_{i=16}^{17} 2^{-i} / 2^{1-2}$; $D(\Delta s_6) = 2^{-4} D(\Delta s_5) + \frac{1}{3}\Sigma_7^2$
$w_2 = s7 = y_1 * s_6$	$\bar{s}_7 := (\bar{y}_1 \cdot \bar{s}_6) \cdot 2^2$	$\mathcal{M}_{y1} = 2^{-1}$; $\mathcal{M}_{s6} = 2^{-1}$; $\lvert s_7 \rvert_{\max} = 1{,}005 < 2^1$; $\mathcal{M}_{s7}(\max) = 2^{-1}$	$\Delta s_7 = \lvert y_1 \rvert_{\max} \Delta s_6 + \lvert s_6 \rvert_{\max} \Delta y_1 + \Sigma_8$ $D(\Delta s_7) = \frac{1}{3}\Sigma^2{}_8 + \lvert y_1 \rvert^2_{\max} D(\Delta s_6)$ $\Sigma_8 = \sum_{i=16}^{29} 2^{-i} / 2^{-1}$

Таблица 3.4. Машинные алгоритмы и масштабные соотношения. Вычисление $w\ (y)$

Операция	Машинный алгоритм	Масштабные соотношения и определение макс. знач.	Расчет погрешности и ее дисперсии (в истинных величинах)
$w_3 = w_1 w_2$	$\overline{w}_3 := (\overline{w}_1 \cdot \overline{w}_2) \cdot 2^2$	$\mathcal{M}_{w1} = 2^{-5}; \mathcal{M}_{w2} = 2^{-1}$ $\lvert w_3 \rvert_{\max} = 18{,}6 < 2^5;$ $\mathcal{M}_{w3}(\max) = 2^{-5}\ ;$ $k = \log_2\left(\dfrac{2^{-5}}{2^{-5} \cdot 2^{-1} \cdot 2^{-1}}\right) = 2$	$\Delta w_3 = \Delta w_1 \lvert w_2 \rvert_{\max} + \Delta w_2 \lvert w_1 \rvert_{\max} + \Sigma;$ $\Sigma = \sum_{i=16}^{29} 2^{-i} / 2^{-5}\ ;$ $D(\Delta w_3) = D(\Delta w_1) \lvert w_2 \rvert^2_{\max} + D(\Delta w_2) \lvert w_1 \rvert^2_{\max} + \frac{1}{3}\Sigma^2$
$w = 2^{-1} w_3$	$\overline{w} := \overline{w}_3$	$\mathcal{M}_w = 2 \cdot \mathcal{M}_{w3} = 2^{-4}$	$\Delta w = 2^{-1} \Delta w_3 ; D(\Delta w) = 2^{-2} D(\Delta w_3)$.

Шаг моделирования выбран равным 64*(1/4096) = 1/64 (на основании соображений, аналогичных рассуждениям в п.1.3)

Ниже приводятся результаты теоретических расчетов, а также экспериментальные значения инструментальной погрешности и ее числовых характеристик.

Таблица 4. Исследование инструментальной погрешности ΔW(y)

Величина	Экспериментальное значение	Теоретическое (расчетное) значение
Максимальная погрешность	0,0015861	0,002733
Средняя погрешность	0,0001630	-
Дисперсия погрешности	3,946E-07	4,738E-07
Среднекв. откл. погрешности	0,0006282	0,0006883

2.3. Фрагмент программы, моделирующий вычислительный блок W(y) СЦВМ (ассемблерная вставка (CPU ix86 ASM) в Delphi Pascal - модуль)

```
function AsmCalcW(Arg1: smallint) : smallint;
const
  const1_20 = $6666;
  const1_6 = $5555;
  const1ofs1 = $2000;
  const1ofs4 = $0400;
  M_Y_FOR_EXP = 4096;
  //for argument interval checking:
  chk_3p_2: smallint = trunc((3*Pi/2)*M_Y_FOR_EXP);
  chk_p_2: smallint = trunc((Pi/2)*M_Y_FOR_EXP);
  chk_m3p_2: smallint = trunc((-3*Pi/2)*M_Y_FOR_EXP);
  chk_mp_2: smallint = trunc((-Pi/2)*M_Y_FOR_EXP);
  chk_5: smallint = trunc(5*M_Y_FOR_EXP);
  chk_m5: smallint = trunc((-5)*M_Y_FOR_EXP);
  //---------------
  const_2p: smallint = trunc(2*Pi*4096); //$6485;
  p_shift_1: smallint = trunc(trunc(Pi*8192)*0.5);
label
  ArgPrepare, CalcSin, Betweenp_2and3p_2,
  Lessthan_p_2, Lessthan_mp_2, Between_m3p_2and_m2p;
var
  W : smallint;
```

```
begin
   asm
     pusha
     pushf
     //calculating exponent...
     mov   bx, arg1
     mov   ax, const1ofs1
     sub   ax, bx  //ax <-- z2, bx <-- y
     imul  bx      //dx:ax
     sal   dx, 1
     sal   ax, 1
     adc   dx, 0  //dx = z3
     mov   ax, const1ofs4
     sub   ax, dx  //ax = z4 = w1
     mov   cx, ax  //w1 = cx
     //calculating sinus argument...  bx = y
     cmp   bx, chk_p_2
     jle   Lessthan_p_2   //otherwise >= pi/2
     cmp   bx, chk_3p_2   //pi/2...3pi/2 ?
     jle   Betweenp_2and3p_2
     sub   bx, const_2p    //here: (3pi/2)...5; y := y - 2*pi
ArgPrepare:
     //argument prepare...
     sal   bx, 2
     jmp   CalcSin
     //other cases...
Betweenp_2and3p_2: //y := pi - y;
     mov   ax, p_shift_1
     sub   ax, bx
     mov   bx, ax
     jmp   ArgPrepare
Between_m3p_2and_m2p:
     add   bx, const_2p
     jmp   ArgPrepare
Lessthan_mp_2:
     cmp   bx, chk_m3p_2
     jl    Between_m3p_2and_m2p
     neg   bx
     sub   bx, p_shift_1
     jmp   ArgPrepare
Lessthan_p_2:
     cmp   bx, chk_mp_2
     jl    Lessthan_mp_2
     jmp   ArgPrepare
     //calculating sinus...
CalcSin:
     mov   ax, bx
     push  bx //store y
     imul  bx {dx:ax <-- s1}
     sal   dx, 1
     sal   ax, 1
     adc   dx, 0  //dx = s1
     push  dx //save s1 to stack
     mov   ax, const1_20
     imul  dx   //dx = s2*2^-2
     sar   dx, 2
     mov   ax, $4000
     sub   ax, dx  //ax = s3
     pop   bx  //retrieve s1 from stack
     imul  bx  //dx:ax <-- s1*s3...
     sal   dx, 1
     sal   ax, 1
     adc   dx, 0
     sal   dx, 1
     sal   ax, 1
     adc   dx, 0 //dx <-- s4
     mov   ax, const1_6
     imul  dx  //dx <-- s5*2^-2
     mov   ax, $4000
```

```
    sub   ax, dx  //ax <-- s6
    pop   bx //restore y
    imul  bx //s7...dx-ax
    sal   dx, 1
    sal   ax, 1
    adc   dx, 0
    sal   dx, 1
    sal   ax, 1
    adc   dx, 0  //dx <-- s7 = w2
    //result
    mov   ax, dx
    imul  cx  //dx:ax <-- w3*2^-2
    sal   dx, 1
    sal   ax, 1
    adc   dx, 0
    sal   dx, 1
    sal   ax, 1
    adc   dx, 0 // result in dx
    //----------------
    mov   w, dx
    //----------------
    popf
    popa
  end;
 Result := W;
end;
```

Выводы

Анализируя результаты, сведенные в табл. 2 и 4 (по каждому из каналов), можно сделать вывод, что теоретические значения инструментальных погрешностей, а также их числовых характеристик (дисперсии и среднеквадратичного отклонения) получаются большими, нежели экспериментальные значения. Это объясняется тем, что в случае аналитических расчетов:

- считается, что возможен крайний случай, когда в отбрасываемых и усекаемых разрядах переменных располагаются одни единицы;

- коэффициенты трансформации равны максимальным значениям соответствующих частных производных.

Иными словами, аналитические расчеты всегда завышают реальную величину погрешности. Вместе с тем заметим, что в рассматриваемых вариантах данное различие (по максимальной ошибке и ее среднеквадратичному отклонению) было не очень большим, это величины одного порядка.

Следовательно, использованный статистический аналитический метод расчета инструментальной погрешности ***может оказаться вполне пригодным для оценочных расчетов погрешности*** при проектировании новых алгоритмов обработки информации в СЦВМ определенной разрядности, а также в процессе разработки микропрограмм.