АНАЛИТИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ АБСОЛЮТНО ТОЧНЫХ ЗНАЧЕНИЙ ОГРАНИЧЕННОГО КОЛИЧЕСТВА СЛУЧАЙНЫХ ЦИФРОВЫХ ДАННЫХ

Х. А. Манаселян

Измерительным процессам и расчетам, включающим в себе компьютерную обработку оцифрованных с помощью АЦП (или накопленных другими способами) цифровых данных, всегда присущи некоторые ошибки. Источниками этих ошибок являются как сами измерительные приборы, чувствительность и стабильность которых не идеальны, так и АЦП, разрядность и быстродействие которых так же бывают ограниченными, хотя эти параметры со временем могут улучшаться. Кроме ошибок начальных данных, ошибки могут возникать и при компьютерной обработке, если надлежащим образом не избавляться от возможных ошибок округлений и интерполяционных или других типов приближений (ошибок усечения).

Естественно, возникает вопрос: а возможно ли вообще полностью избавиться от ошибок, связанных с компьютерной обработкой? Когда речь идет об абсолютной точности, то необходимо сделать некоторые предварительные оговорки. Во-первых, с выхода АЦП поступают в компьютер уже округленные значения чисел, и от этого округления никак нельзя избавиться из-за неизбежных технических ограничений. Во вторых, практически все известные методы приближения, которые применяются в цифровых обработках сигналов [1, 2], для того, чтобы ошибки усечения сводились к минимуму (часто, не к нулю, так как, например, при случае нестационарного случайного процесса этому достичь невозможно теоретически), ставится требование, чтобы параметры приближения (количество членов сумм рядов, итерационных циклов и т.д.) стремились к бесконечности. А для компьютерных программ, из-за ограниченного объема оперативной памяти, практически это является таким же запрещенным приемом, как деление на нуль. С другой стороны, подавляющее большинство известных нам математических программ вводят так же и ошибки округления, которые, накапливаясь и возрастая, могут сводить на нет результаты применения любой высокоточной методики приближения. Тогда о какой абсолютной точности может идти речь?

Чтобы достаточно корректно ответить на поставленный вопрос, в данной работе речь будет идти о массиве цифровых данных, которые представляют из себя либо оцифрованные с какой-то точностью значения аналогового сигнала, либо являются просто набором конечных рациональных чисел, полученных простым (не автоматическим) накоплением. Будем предполагать, что значения чисел даны нам абсолютно точно, то есть здесь мы не будем рассматривать ошибки, которые возникают в самих измерительных приборах и в АЦП. В качестве рациональных чисел, в частном случае, могут быть и целые или натуральные числа, например, конечный набор n простых чисел, идущих по порядку возрастания или неупорядоченные по величине.

Как процедуру индексации, так и дальнейшие действия компьютерной обработки будем осуществлять с помощью программы “Mathematica 6”, в среде которой, как покажем далее на конкретном примере, практически полностью можно избавиться от ошибок округления, задавая точность представления чисел с достаточным запасом. При этом можно исходить из максимального количества тех знаков после запятой, которые присутствуют в членах массива данных. Рассмотрим упрощенный случай, когда имеется всего один массив, а в операционной среде Windows XP запущен только программа “Mathematica 6”. Допустим, что пользуемся персональным компьютером, объем оперативной памяти которого составляет 2 GB, что вполне реальная цифра для современных компьютеров средней мощности. Оценим максимально возможный размер одного массива данных, при котором оперативная память компьютера может перегружаться, заметно замедляя работу программы “Mathematica 6”. Для нормальной работы этой программы необходимо иметь не менее 50 MB свободной оперативной

памяти, а сама операционная система Windows XP занимает обычно не более 250 MB. Следовательно, при этих условиях у нас остается 1,7 GB свободной памяти, поэтому для максимального количества чисел, каждый из которых занимает, скажем, 16 B памяти, получаем ориентировочную оценку $n_{\max} \approx 1.0625 \times 10^8$.

Если, допустим, количество чисел в массиве меньше этого числа на три порядка, то точность представления чисел мы искусственно можем увеличить, изменяя точность их представления так, чтобы каждое из них занимало объем памяти, например, 160 B вместо 16-ти (программа "Mathematica 6" дает такую возможность, благодаря достаточно богатому арсеналу встроенных функций представления и форматирования чисел). С учетом также того, что данная программа во время своей работы своевременно стирает промежуточные ненужные звеня расчетов, то в большинстве случаев такое форматирование чисел уже позволяет полностью избавиться от всевозможных ошибок округления, не перегружая при этом оперативную память. Иными словами, если в среде какой-либо математической программы точность представления чисел не имеет существенных ограничений, то ошибки округления могут стать неизбежными только тогда, когда объем оперативной памяти компьютера не достаточно большой.

Приступим сейчас к обсуждению вопроса, связанного с ошибками, возникающими при интерполировании цифровых данных, например, с помощью конечной суммы степенного ряда (многочлена n-го порядка).

Предварительно скажем, что если скорость оцифровки АЦП настолько большая, что в течение каждой секунды на компьютер подаются огромные массивы данных, то для их обработки в реальном времени, конечно же, наилучшим приближением можно считать гармонический анализ с помощью рядов Фурье, например, БПФ. По существу, преимуществом гармонического анализа заключается в том, что коэффициенты a_n и b_n, стоящие у тригонометрических функций членов ряда, сравнительно быстро стремятся к нулю, когда n стремится к бесконечности. Это обстоятельство в свою очередь приводит к тому, что среднеквадратическая ошибка приближения, начиная с какого то определенного значения n, становится наименьшей, по сравнению с другими ортогональными или ортонормированными базисами приближения. Однако на практике не всегда перед операторами стоят столь сложные задачи, вроде как предельно точная функциональная аппроксимация происходящего физического процесса во временной или спектральной областях (или в обоих областях одновременно) с использованием огромных массивов данных, предварительно уплотненных в АЦП или в самом компьютере. Нередко цифровые данные бывают не столь большими, и поэтому нас сейчас интересует следующий вопрос – может ли оператор в таких случаях найти альтернативное приближение, которое обеспечивает более высокую точность приближения по сравнению с приближением с помощью конечной суммы тригонометрического ряда. Более того, мы ставим перед оператором еще более «тенденциозную» задачу, а именно, найти и применить аналитическое представление, обеспечивающее абсолютную точность при обратном расчете с помощью этого представления всех без исключения значений чисел данного массива. И еще, ставим требование ограничиваться конечными суммами рядов, у которых имеется членов не больше, чем имеется чисел в массиве. Не трудно догадаться, что если оператору это удастся сделать, то вместо интерполяционной формулы он получит просто формулу общего члена последовательности n чисел, которые в общем случае могут быть некоррелированными между собой. Возможно ли это?

Первым делом, как это обычно делается, будем приписывать каждому члену линейного массива данных свой определенный индекс. Мы здесь намерены показать, что этого будет достаточно, чтобы, используя эти индексы в качестве множества аргументов для вполне определенного многочлена, получить все значения членов данного набора чисел с помощью этого многочлена. По существу, это не что иное, как обратная задача интерполяции массива данных, исходя из конечного числа значений степенных (или других, не обязательно ортогональных) функций, коэффициенты при которых заранее не

известны и прямым образом зависят от величин и количества самих чисел данного массива.

Во избежание того, чтобы дальнейшие выкладки не получились громоздкими, ограничимся одномерным массивом, имеющем всего 10 членов. Воспользуемся генератором случайных чисел **RandomReal** и ограничимся десятизначной точностью представления чисел:

```
a=RandomReal[{-100, 100}, 10];
SetPrecision[a, 10]
{63.11513075, -42.21916849, 30.27566028, 37.48778083, 26.83175681, 75.60555908,
 -20.06853545, 90.64613215, -57.80905185, -39.33441336}
```

Превратим полученные дробные числа в целые, умножая каждое из них на 10^8:

```
c=Round[a*10^8]
{6311513075, -4221916849, -3027566028, 3748778083, 2683175681, 7560555908, -2006853545,
9064613215, -5780905185, -3933441336}
```

Далее, интерполируем массив «c», воспользовавшись встроенной функцией **InterpolatingPolynomial**:

```
InterpolatingPolynomial[c,x]
6311513075+(-10533429924+(11727780745/2+(-6145787455/6+(-1819538087/6+(34487021293/120+(-
95868591689/720+(17635307267/360+(-156168779227/10080+(307425471743 (-9+x))/72576) (-8+x)) (-7+x))
(-6+x)) (-5+x)) (-4+x)) (-3+x)) (-2+x)) (-1+x)
```

Нам остается доказать, что примененная нами интерполяция абсолютно точна, т.е. с помощью полученного выражения можно восстановить значения чисел первоначального массива «a». Для этой цели применим простой цикл **For**, пробегающий индексы от 1 до 10, разделив каждый раз результаты расчета по полученному нами выражению на 10^8, чтобы они стали десятичными дробями, как это было в начале:

```
b={};
For[i=1,i<=10,A=N[(6311513075+(-10533429924+(11727780745/2+(-6145787455/6+
(-1819538087/6+(34487021293/120+(-95868591689/720+(17635307267/360+
(-156168779227/10080+(307425471743 (-9+i))/72576) (-8+i)) (-7+i)) (-6+i)) (-5+i)) (-4+i)) (-3+i)) (-2+i))
(-1+i))/10^8,10];AppendTo[b,A];i++];Print[b]
{63.11513075, -42.21916849, -30.27566028, 37.48778083, 26.83175681, 75.60555908, -20.06853545,
90.64613215, -57.80905185, -39.33441336}
```

Сравнивая числа в таблицах «a» и «b», действительно убеждаемся, что нами получены абсолютно точные значения первоначальных данных. Это, во первых, говорит о том, что в среде компьютерной программы “Mathematica 6” можно осуществить абсолютно точную интерполяцию ограниченного количества чисел, в виде аддитивных многочленов с целыми коэффициентами, стоящими у степеней индексов данных чисел.

Но на первый взгляд может показаться, что это всего лишь заслуга встроенной функции **InterpolatingPolynomial**, и, что наряду с увеличением разрядности или количества чисел достигнутая точность постепенно будет ухудшаться. Однако это не так. По существу, данный вид интерполяции равносилен нахождению аналитического решения системы из n линейных уравнений с n неизвестными. Действительно, если степени индексов мы рассмотрим как известные коэффициенты, а искомые коэффициенты, стоящих у каждого из них, как неизвестные, то легко убедиться, что получим именно такую систему уравнений, решение которой можно представить абсолютно точно, записав найденные значения неизвестных в виде обыкновенных дробей.

Существенный недостаток данного подхода заключается в том, что если увеличим количество чисел n, скажем, на единицу, то все найденные коэффициенты будут

меняться. Все же описанный выше подход может успешно применяться в тех случаях, если все члены массива являются целыми числами и, при компьютерных манипуляциях с ними, ошибки, возникающие при переходе на десятичные дроби, являются нежелательными.

В качестве примера рассмотрим простой случай, когда массив содержит всего 10 простых чисел, идущих по порядку возрастания:

11, 13, 17, 19, 23, 29, 31, 37, 41, 43;

Известно, что трудно найти общую аналитическую формулу, зависящую от одной переменной, чтобы, пробегая все 10 значения этой переменной, получить все заданные простые числа. Примером такой формулы может служить известный квадратичный трехчлен Эйлера $x^2 + x + 41$, который для натуральных значений переменной от 1 до 39 дает разные простые числа. Мы же здесь ставим обратную задачу, а именно, пробуем получить формулу в виде многочлена, зависящую от порядковых номеров заданных простых чисел.

Неявный вид искомого многочлена запишем следующим образом:

$$P_i = \sum_{k=0}^{9} A_k i^k \qquad (1),$$

где A_k некие, постоянные для всех P_i, коэффициенты. Совершим следующую замену переменных: рассчитывая значения всех i^k, обозначим их через $a_{m,n}$, где m пробегает значения i, а n соответствует значениям k, а так как значения A_k нам пока не известны, то обозначив их через x_i. Тогда на основании (1) получим следующую систему уравнений:

$$\left\{\begin{array}{l} P_1 = a_{1,1}x_1 + a_{1,2}x_2 + ... + a_{1,10}x_{10} \\ P_2 = a_{2,1}x_1 + a_{2,2}x_2 + ... + a_{2,10}x_{10} \\ ... \\ P_{10} = a_{10,1}x_1 + a_{10,2}x_2 + ... + a_{10,10}x_{10} \end{array}\right\} \qquad (2).$$

Решив систему (2) и поставив соответствующие значения переменных $x_1, ..., x_{10}$ вместо коэффициентов $A_1, ..., A_{10}$ в (1), окончательно получим:

$$P_i = -297i^0 + \frac{221423}{252}i^1 - \frac{1019315}{1008}i^2 + \frac{28342031}{45360}i^3 - \frac{662747}{2880}i^4 +$$
$$+ \frac{114139}{2160}i^5 - \frac{2195}{288}i^6 + \frac{10133}{15120}i^7 - \frac{661}{20160}i^8 + \frac{31}{45360}i^9 \qquad (3).$$

Умножим многочлен (3) на НОК знаменателей дробей $A_1, ..., A_{10}$. В результате получим многочлен с целыми коэффициентами, что позволит нам избавиться от ошибок округления, которые могут возникать при превращении обыкновенных несократимых дробей в десятичные:

$$181440P_i = -53887680i^0 + 159424560i^1 - 183476700i^2 + 113368124i^3 -$$
$$-41753061i^4 + 9587676i^5 - 1382850i^6 + 121596i^7 - 5949i^8 + 124i^9 \qquad (4).$$

Для $i = 1$ из (4), получим:

$$P_1 = \frac{1995840}{181440} = 11.$$

Легко убедиться, что подобным же образом можно получить и все остальные простые числа данной нам последовательности, т.е. в данном случае ошибки округления не возникнут нигде, в результате чего первоначальные значения чисел мы получим с абсолютной точностью. Для этого достаточно просто рассчитать их значения с помощью

многочлена (4). Иными словами, выражение (3) представляет собой формулу общего члена данной нам последовательности из 10-ти простых чисел. Задача решена.

Возникает вопрос: Во всех ли случаях нам придется умножать многочлен типа (3) на НОК знаменателей дробей $A_1,...,A_n$, чтобы коэффициенты этого многочлена стали целыми?

Чтобы ответить на этот вопрос, оформим две отдельные теоремы:

1) *Любую пару простых чисел можно представить в виде многочлена* $P_x = Ax + B$, *где* $x \in \{1;2\}$, A *и* B *– целые числа;*

2) *Любую тройку нечетных простых чисел можно представить в виде многочлена* $P_x = Ax^2 + Bx + C$, *где* $x \in \{1;2;3\}$, A, B *и* C *– целые числа.*

Докажем вторую теорему (доказательство первой производится аналогичным образом). Составим систему уравнений для значений x аналогично (2) и выразим коэффициенты A, B и C через P_1, P_2, P_3:

$$A = \frac{P_1 - 2P_2 + P_3}{2};\ B = \frac{-5P_1 + 8P_2 - 3P_3}{2};\ C = 3(P_1 - P_2) + P_3.$$

Так как P_1, P_2 и P_3 являются нечетными числами, то числители дробей A и B будут четными, и дроби сократятся на 2. Теорема доказана.

Резюмируя все вышеизложенное, можно привести два важных заключения. Если ошибки округления полностью отсутствуют или используемая компьютерная программа позволяет избавиться от них, то:

1) *члены любого ограниченного одномерного массива, которые могут иметь случайные и независимые друг от друга величины, можно представить одним общим многочленом степени n-1 с абсолютно точными значениями коэффициентов, где n есть количество чисел данного массива*;

2) *для любой последовательности n простых чисел можно найти формулу, являющуюся общим членом для данной последовательности.*

Обсудим второе заключение более подробно. Естественно, что рассмотренная нами задача имеет решение и тогда, когда n есть произвольное натуральное число, значительно превышающее число 10, которое мы взяли в качестве примера. Все же, второе заключение может выглядеть неправдоподобным, хотя оно верно, как мы это показали на конкретном примере. Вместе с тем, верно так же и то, что не существует многочлена $f(x)$ от переменной x с целыми коэффициентами, который для всех натуральных значений этой переменной дает простое число. Данное утверждение доказано во многих книгах по простым числам (например, в [3]). Однако оно не входит в противоречие с нашим заключением, т.к. в нашем случае переменная пробегает ограниченное количество значений, которое совпадает с количеством простых чисел, входящих в данную нам последовательность.

Какая же может быть польза от данного подхода представления простых чисел?
Ответ на этот вопрос следующий. Известно, что в настоящее время многие математики мира заняты тем, что ищут очередное гигантское простое число типа простых чисел Мерсена. Недавно было сообщение о том, что калифорнийским математикам удалось с помощью совместной работы 75-ти мощных компьютеров найти очередное (45-ое) простое число Мерсена, которое содержит свыше 13 миллионов знаков. Известно также, что предыдущее такое число было найдено целых два года тому назад, но количество его знаков, к сожалению, не пересекло границу 10 миллионов, за что была предназначена денежная премия.

Поиск новых, достаточно больших простых чисел имеет не только чисто познавательное значение. Простые числа широко применяются, например, в криптографии [4, 5]. Логично предполагать, что чем больше ключей заложено в зашифрованный код, тем труднее его расшифровать. Именно в этом аспекте описанный

здесь метод представления простых чисел может внести положительный вклад. Например, если в качестве ключа используются простые числа Мерсена, общая формула которых имеет вид $P_M = 2^p - 1$, где p некое простое число, то количество ключей можно существенно увеличить, если последовательность нескольких простых чисел Мерсена представить с помощью одного многочлена с целыми коэффициентами. Можно идти дальше в этом направлении и использовать не простые числа Мерсена, а простые числа более общего вида, по отношении которых числа Мерсена являются частным случаем. Такими числами являются простые числа вида $P_a = \frac{a^p - 1}{a - 1}$, где a любое целое число, а $p-$ некое простое число. Нетрудно убедиться, что простые числа этого вида имеют те же свойства, как и числа Мерсена, общий вид которых получим, если поставим в приведенной формуле $a = 2$. При значении же $a = -2$ получим известные числа $P_W = \frac{2^p + 1}{3}$, именуемые в литературе как простые числа Вагстафа [6].

В таблице 1 приведены некоторые произвольно выбранные простые числа типа P_a, где даны как значения a и p, так и количество знаков соответствующих простых чисел.

Таблица1

$\pm a$	p	колич. знаков	$\pm a$	p	колич. знаков
39	631	1003	715	1531	4368
-42	709	1150	2057	1669	5527
57	661	1159	-4123	1747	6313
-30	829	1224	-4256	1747	6337
-60	937	1665	-5359	1789	6668
-54	991	1716	8787	1811	7139
-84	971	1867	13299	1949	8034
-209	883	2047	-22089	1973	8567
-84	1087	2090	-27183	1993	8834
133	991	2103	31906	1999	8999
-105	1187	2398	39529	2011	9240
133	1181	2507	-49290	2039	9564
-154	1217	2661	-57289	2153	10240

Так как для простых чисел типа P_a в качестве основания a может служить любое целое число, то ясно, что в любом числовом интервале натурального ряда простые числа этого типа можно встретить несравненно чаще, чем простые числа Мерсена или Вагстафа. Данный факт можно рассматривать как определенное преимущество, так как простые числа этого типа могут обеспечивать более широкое поле для их отбора. Это важно, например, для систем шифрования «с открытым ключом» (таких, как RSA и др.), когда из соображения безопасности и стойкости зашифровки периодически приходится выбирать близкие по количеству знаков различные пары простых чисел.

Есть еще один интересный момент, на котором здесь следует остановиться. Известно, что для простых чисел Мерсена существует детерминистический код проверки простоты, который опирается на следующую теорему:

Теорема Люка-Лемера: *Число* $P_M = 2^p - 1$, *где* $p-$ *простое нечетное число, тогда и только тогда есть простое, когда* P_M *является делителем* $(p-2)$*-го члена последовательности* u_n *, определенный условиями* $u_0 = 4,\ u_{n+1} = u_n^{\,2} - 2,\ n = 0,1,2,\ldots$.

Из этой теоремы следует, что $P_M \mid p-2$ тогда и только тогда, когда P_M является делителем $(p-2)$-го члена последовательности $r_n\,(n=0,1,2,...)$, зависящей от P_M и определенной условиями: $r_0=4$, r_{n+1} является остатком от деления r_n^2-2 на P_M.

Соответствующий код в среде "Mathematica 6" выглядит следующим образом:

```
LucasLehmer[p_]:=NestList[Mod[#^2-2, 2^p-1]&, 4, p-2]
```

Хотя есть возможность предпринимать определенные шаги в направление того, чтобы тест Люка-Лемера работал быстрее, но вряд ли можно считать оправданным то, что внимание многих групп математиков разных стран сосредоточено в основном на поиски простых чисел Мерсена, которые встречаются все реже и реже, по мере увеличения размеров чисел. Помимо этого, достоверно известно, что в любом числовом интервале натуральных чисел, имеющих одинаковое количество знаков, существуют как минимум три простых числа. Мы здесь к этому добавим и то, что все эти три простые числа можно представить в виде P_a. Чтобы не быть голословным, докажем следующую теорему:

6) *Для любого нечетного числа P (в том числе – и простого) существуют хотя бы одно целое число a и одно простое число $p, (p<P)$, удовлетворяющие формуле* $P=\dfrac{a^p-1}{a-1}$.

Доказательство этой теоремы станет очевидным, если для заданного нечетного (или простого) числа P в качестве искомого значения a брать $P-1$, а в качестве $p-$первое простое число. Тогда для правой части приведенной в теореме формулы получим:

$$\frac{a^p-1}{a-1}=\frac{a^2-1}{a-1}=a+1\equiv P.$$

Очевидно, что сделанный нами выбор для значений a и p может быть не единственным. Например, для чисел Мерсена, помимо указанных нами значений a и p, приведенной в теореме формуле удовлетворяют также значение $a=2$ и некое простое число p, в качестве показателя степени двойки. Поэтому, для этого типа чисел аналогичное вышеприведенной теореме 6 утверждение принимает несколько шутливую форму:

7) *Некоторые нечетные числа (в их числе встречаются и простые) можно представить в виде $P=2^p-1$, где p есть некое простое число.*

Иногда дается предпочтение простым числам Мерсена именно за то, что они довольно редко встречаются в натуральном ряде. Однако заметим, что если вместо двойки зафиксировать любое другое значение целого числа a, то простые числа типа P_a в натуральном ряде будут встречаться так же редко, как и простые числа Мерсена. Если же задача поставлена так, что нужно выбирать простое число, имеющее определенное количество знаков в своей записи, то этого легче достичь варьированием значения a, так как в заданном числовом интервале может не оказаться ни одного простого числа Мерсена. Например, начиная с сороказначных десятичных чисел и кончая 156-зачными нет ни одного простого числа Мерсена, в то время как, например, $\dfrac{3^{281}+1}{4}$ есть простое число $(a=-3)$, имеющее в своей записи 134 цифровых знаков.

Литература

1. G. A. Korn, T. M. Korn, *Mathematical Handbook for Scientists and Engineers,* 2nd Enlarged and Revised Edition, McGraw-Hill Book Company, New York San Francisco Toronto London Sydney, 1968.

2. George W. Collins, II, *Fundamental Numerical Methods and Data Analysis,* Internet Edition, 2003.
3. В. Серпинский, *Что мы знаем и чего не знаем о простых числах,* Гос. Изд. физмат литературы, Москва, 1963.
4. Н. Коблиц, *Курс теории чисел и криптографии,* Москва, Науч. Изд. ТВП, 2001.
5. Н. Фергюсон, Б. Шнайер, *Практическая криптография,* Изд. «Диалектика-Вильямс», 2004.
6. Bateman, P. T., Selfridge, J. L., and Wagstaff, S. S. *The New Mersenne Conjecture*, Amer. Math. Monthly **96**, 125-128, 1989.