

Оптимизация работы с числами расширенной точности в программном комплексе Global Expert

В.Г. Чигарев, В.В. Рябов

Нижегородский государственный университет им. Н.И. Лобачевского

Введение

На кафедре математического обеспечения ЭВМ факультета ВМК Нижегородского государственного университета им. Н.И. Лобачевского разрабатывается программный комплекс Global Expert, предназначенный для параллельного решения задач многомерной многоэкстремальной оптимизации на кластерных системах.

В данной работе рассматривается оптимизация работы с числами расширенной точности, заключающаяся в универсализации интерфейса для использования готовых библиотек чисел расширенной точности, внедрение одной из самых быстрых библиотек для работы с числами расширенной точности - MPFR в систему Global Expert и тестирование модуля чисел расширенной точности.

Числа расширенной точности

Под числом расширенной точности имеют в виду структуру или класс, которые позволяют выполнять арифметические действия с произвольной точностью.

С ростом размерности решаемых задач и степени кривой Пеано вся область поиска сужается до отрезка от 0 до $L+1$, где L – количество дополнительных разверток. Это означает, что и точность, необходимая для представления образов точек на отрезке, а так же расстояний между ними и достижение требуемой точности вычислений становится все более и более трудной задачей. При этом возникает ряд проблем, одной из которых является нехватка мантиссы стандартного типа double для хранения координаты точки испытания. Данный факт можно проиллюстрировать на примере:

```
double x1 = 0.22222222222223; // 14 знаков  
x1 /= 2;
```

Фактически, после деления в числе $x1$ должно находиться 0.11111111111115 (15 знаков), а будет находиться 0.11111111111112 (14 знаков)

Если в качестве типа данных для хранения координаты точки испытания использовать double, то максимальной областью применимости системы Global Expert являются все задачи, в которых $n*m < 52$, что серьезно ограничивает возможности комплекса. Условие $n*m < 52$ появляется из-за того, что в типе double отводится 52 бита на хранение мантиссы, поэтому максимальное число различных точек не превышает 2^{52} .

В связи с чем, возникает острая необходимость использования чисел расширенной точности для решения задач высокой размерности.

Использование в предыдущих версиях Global Expert

Изначально в рамках комплекса для представления чисел расширенной точности использовалась библиотека MAPM. Но в результате сравнения скоростей операций над числами в библиотеках MAPM и MPFR оказалось, что последняя производит часто используемые комплексом Global Expert операции быстрее, чем первая. Таким образом, появилась идея реорганизовать код таким образом, чтобы можно было легко интегрировать библиотеки для работы с числами расширенной точности и добавить возможность использования библиотеки MPFR в систему.

Результаты сравнения скоростей выполнения операций в библиотеках MAPM и MPFR приведены в таблице:

Операция	Количество операций	Время выполнения, сек.	
		MAPM	MPFR
+	1000000	6.703	7.828
*	1000000	14.640	8.610
/	1000000	7.094	7.703
√	100000	78.062	49.985

Также в ходе исследования внутреннего представления чисел в библиотеках оказалось, что для представления одного и того же числа одинаковой точности в библиотеке MPFR требуется меньше объема памяти, чем в библиотеке MAPM.

Данные о необходимых объемах памяти используемых для представления одного числа с точностью $1e(-N)$ в разных библиотеках приведены в таблице:

Библиотека	Структура, байт	Данные, байт	Всего, байт
MAPM	28	4N	28 + 4N
MPFR	16	$N / \lg 2$	16 + 3,32N

Таким образом, общая разница составляет: $12 + 0,68 \cdot N$ бит для каждого числа, что должно положительно отразиться на объеме используемой памяти при выборе библиотеки MPFR.

Результаты сравнения используемой памяти и скорости работы алгоритма для библиотек чисел расширенной точности.

Решалась задача Растригина со следующими параметрами:

- N - 6
- метод - последовательный с одной разверткой
- $\epsilon_{ps} - 0.01$
- m - 8
- $r - 2.5$

	MAPM	MPFR
Время счета, сек	80.091688	42.736168
Используемая память, Кбайт	13924	10928

Найденный минимум в обоих случаях совпадает.

Библиотека MAPM

MAPM расшифровывается как Mike's Arbitrary Precision Math Library. Идеи реализации библиотеки были заимствованы ее автором из библиотеки APM 1988 года. Отличительные особенности библиотеки MAPM:

- 1) округление до произвольной точности;
- 2) наличие тригонометрических и других математических функций.

Внутреннее представление числа является числом в 100-ичной системе счисления. Каждый байт данных может хранить значение от 0 до 99 (вместо возможных 0-255). Основание равно степени 10 было выбрано в силу того, что такие числа удобно переводить в/из внутреннего представления для представления, привычного человеку.

Библиотека MPFR

MPFR – портируемая, свободно распространяемая библиотека написанная на C для выполнения операций над числами с плавающей запятой до любого порядка точности. Она базируется на библиотеке GNU MP. Основные особенности:

- 1) код библиотеки портируемый (не зависит от размера машинного слова);
- 2) возможность изменения точности для каждого числа;
- 3) предоставляет 4 способа округления из стандарта IEEE 754-1985.

Внутреннее представление числа в данной библиотеке следующее:

$$\text{sign} * (d[k-1]/B + d[k-2]/B^2 + \dots + d[0]/B^k) * 2^{\text{exp}},$$

где $k = \text{ceil}(\text{mp_prec}/\text{BITS_PER_MP_LIMB})$, $B = 2^{\text{BITS_PER_MP_LIMB}}$, d – массив данных.

Реализация класса расширенной точности

Основная часть реализации сосредоточена в классе `Extended`. Для универсализации использования библиотек чисел расширенной точности в классе `Extended` используются указатели на объекты, представляющие числа в разных библиотеках. Такой подход позволяет добавить новую библиотеку для чисел расширенной точности простым добавлением нового указателя и описанием математических операций. Таким образом, класс `Extended` является оберткой для используемых библиотек.

В реализации библиотек `MAPM` и `MPFR` число расширенной точности представляет собой объект класса `C++`. Поэтому, для организации межпроцессорного взаимодействия, а так же для обеспечения возможности записи чисел расширенной точности на диск были разработаны методы упаковки и распаковки, позволяющие переводить числа расширенной точности в непрерывную область памяти.

Для выбора используемой библиотеки для представления чисел расширенной точности были добавлены опции командной строки в подсистему оптимизации, которые позволяют задать используемую библиотеку для чисел расширенной точности (`MPFR`, `MAPM` или `double`) и точность. По умолчанию точность устанавливается в $2^{((m+1)*N)}$, где m – порядок кривой Пеано, а N – размерность задачи. $2^{((m+1)*N)}$ – максимальное количество точек, которое может быть сгенерировано алгоритмом. Использование точности, меньшей, чем это число не может гарантировать отсутствие слияния двух соседних точек.

Модульное тестирование

На данный момент библиотек модульного тестирования для кода, написанного на языке `C++` достаточно много. Наиболее популярная из них это библиотека `CppUnit`. Однако в нашем случае более подходящей оказалась библиотека `unit--`, которая сочетает в себе простоту, легкость в использовании и достаточную для нас гибкость. Она позволяет достаточно просто описывать тесты и объединять их в группы при помощи специальных макросов.

Написание тестов способствовало улучшению качества кода, уменьшению количества ошибок. В дальнейшем, при изменении реализации класса чисел расширенной точности можно будет использовать регрессионный тест для проверки работоспособности основной функциональности модуля после внесенных изменений.

Литература

1. Стронгин Р.Г. Численные методы в многоэкстремальных задачах. (Информационно-статистические алгоритмы). М., Наука, 1978.
2. Орлянская И.В. Современные подходы к построению методов глобальной оптимизации. Электронный журнал «Исследовано в России», 2007г. МГУ, факультет ВМиК.
3. Рябов В.В., Сысоев А.В. Алгоритмы принятия глобально-оптимальных решений и их модификации. // Материалы международной конференции «Высокопроизводительные вычисления на кластерных системах», Нижний Новгород, 2006.
4. Баркалов К.А., Рябов В.В., Сидоров С.В., Сысоев А.В. Об опыте решения задач многоэкстремальной оптимизации на высокопроизводительных кластерных системах. // Материалы конференции «Технологии Microsoft в теории и практике программирования - 2008».
5. Информационный источник <http://www.mpfr.org>.
6. Информационный источник <http://www.opensourcetesting.org>.