

**Экономичный алгоритм поточной
свертки в проблеме
вибросейсмоакустического
мониторинга.**

**Хайретдинов М.С., Седухина Г.Ф., Воскобойникова Г.М.,
Матвеев И.Н.**

**Институт вычислительной математики и математической
геофизики СО РАН**

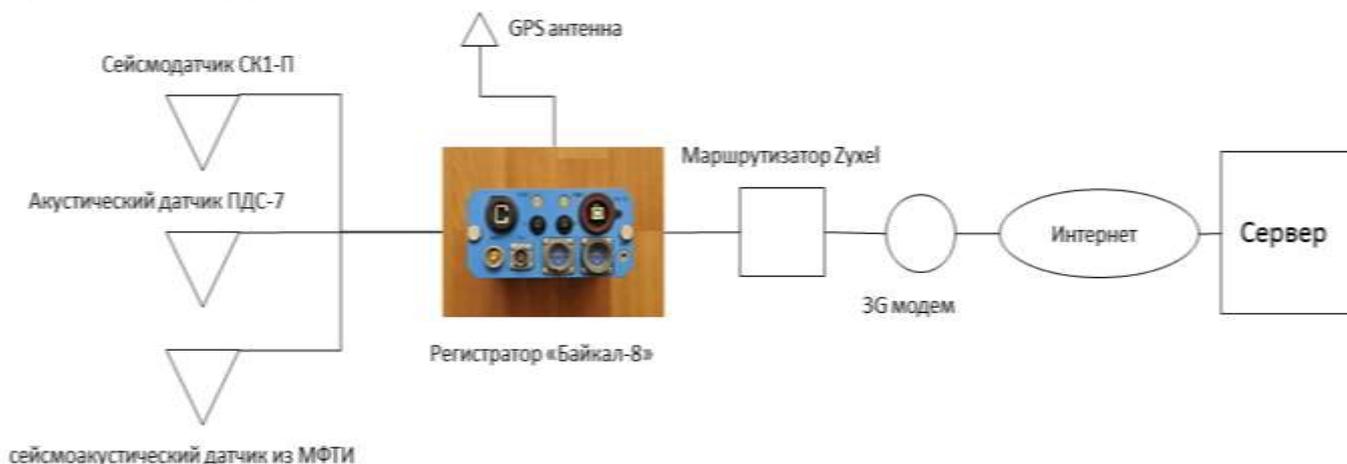
Тринадцатая международная азиатская школа-семинар

«Проблемы оптимизации сложных систем»

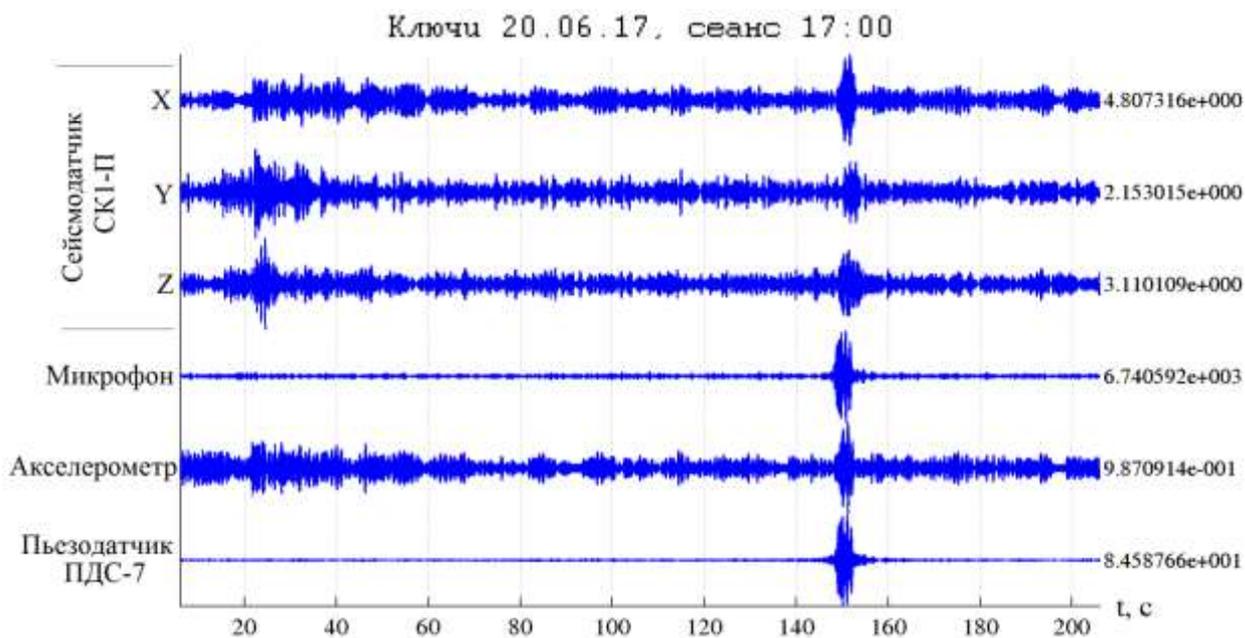
18-22 сентября 2017 г.

Новосибирск

Введение. Внедрение сетевых технологий сбора данных в системы активного геофизического мониторинга окружающей среды [1] открывает возможности оперативного анализа данных в темпе их поступления. Это обуславливает возможности поточной обработки данных в системах активного мониторинга с применением вибрационных сейсмических [2] и акустических [3] колебаний.



Как известно, одной из основных используемых операций здесь является взаимокорреляционная свертка продолжительных по времени регистрируемых и опорных сигналов, в результате которой получается коррелотрасса.



Вычисление этой операции в реальном масштабе времени, т.е. в темпе поступления данных при априорно неизвестных временах прихода сейсмических и акустических колебаний, возможно лишь путем секционирования длинных временных последовательностей. К тому же,

использование алгоритма быстрого преобразования Фурье (БПФ) для обработки секционированных данных обеспечивает возможности высокопроизводительных вычислений.

Пусть Y – искомая коррелотрасса. При секционировании оценка ее вычисляется в итерационном виде: $Y^{i+1} = f(Y^i, X_{i+1})$, где Y^i – приближение Y на i -м шаге, X_{i+1} – текущие отсчеты данных.

Секционированные входные данные по мере поступления накапливаются во вспомогательных буферах и после заполнения очередного буфера обрабатываются при помощи БПФ. Таким образом, для того, чтобы иметь возможность обрабатывать сигналы в режиме реального времени с использованием БПФ, надо представить аperiodическую корреляционную свертку двух массивов большого размера как сумму циклических обычных свертки частей (секций) этих массивов. Причем это разбиение надо сделать таким образом, чтобы трудоемкость обработки одной секции зависела только от величины этой секции, и не зависела от множества всех входных отсчетов размера N . Это аналогично тому, что имеет место в цифровой фильтрации, когда размер опорной последовательности, представляющей дискретный аналог импульсной функции фильтра, несоизмеримо меньше размеров входных и выходных массивов, которые можно считать бесконечными. Напротив, в задачах обработки вибросейсмических сигналов входной X и опорный S массивы имеют одинаково большие размерности, а результат их свертки Y представляется сравнительно небольшим числом отсчетов M ($M \ll N$). Это обстоятельство лежит в основе выбора подхода к разбиению массивов X и S . С учетом этого алгоритм свертки обоих массивов может быть представлен в виде

$$Y_m = \sum_{n=0}^{N-1} X_n \cdot S_{n-m} = \sum_{l=0}^{\frac{N}{M}-1} \sum_{n=0}^{M-1} X_{M \cdot l + n} \cdot S_{M \cdot l + n - m} = \sum_{l=0}^{\frac{N}{M}-1} \sum_{n=0}^{2L-1} A_n^l \cdot B_{\langle m-n \rangle_{2M}}^l, \quad (1)$$

где M – размер одной секции входного массива (подразумевается, что N делится на M нацело. В противном случае массив расширяется путем дописывания нулей). A и B – массивы, состоящие из $2L$ отсчетов и определяемые следующим образом:

$$A_i^l = \begin{cases} X_{M \cdot l + i}, & \text{если } 0 \leq i < M \\ 0, & \text{если } M \leq i < 2M \end{cases}$$

$$B_i^l = S_{M \cdot (l-1) + \langle M - i \rangle_{2M}}$$

$$0 \leq i < 2M$$

В соответствии с теоремой о свертке для дискретного случая, циклическая свертка двух массивов равна обратному дискретному преобразованию Фурье (ОДПФ) от произведения прямых ДПФ этих массивов, следовательно:

$$Y_m = \frac{1}{N} \sum_{l=0}^{M-1} \sum_{k=0}^{2M-1} \left(\sum_{n=0}^{2M-1} A_n^l \cdot W^{-n \cdot k} \right) \cdot \left(\sum_{l=0}^{2M-1} B_n^l \cdot W^{-l \cdot k} \right) \cdot W^{k \cdot m},$$

где $W = e^{j \cdot \frac{2\pi}{2M}}$

На рис. 1 приведена схема, иллюстрирующая порядок действий, производимых при обработке i -ой секции входной последовательности X . Значок * означает БПФ от массивов и умножение спектра массива A на спектр, комплексно сопряженный со спектром массива B . Такие действия производятся для каждой секции входной последовательности. По завершении работы вычисляется коррелограмма путем выполнения обратного БПФ от спектра Y .

Трудоемкость используемого алгоритма БПФ (двухточечного, рассчитанного на вещественные данные) составляет $\frac{N}{2} \cdot (3 \cdot \log_2 N - 5) + 4$ операций сложения и $N \cdot (\log_2 N - 3) + 4$ операций умножения, т.е. в сумме:

$$\frac{N}{2} \cdot (5 \cdot \log_2 N - 11) + 8.$$

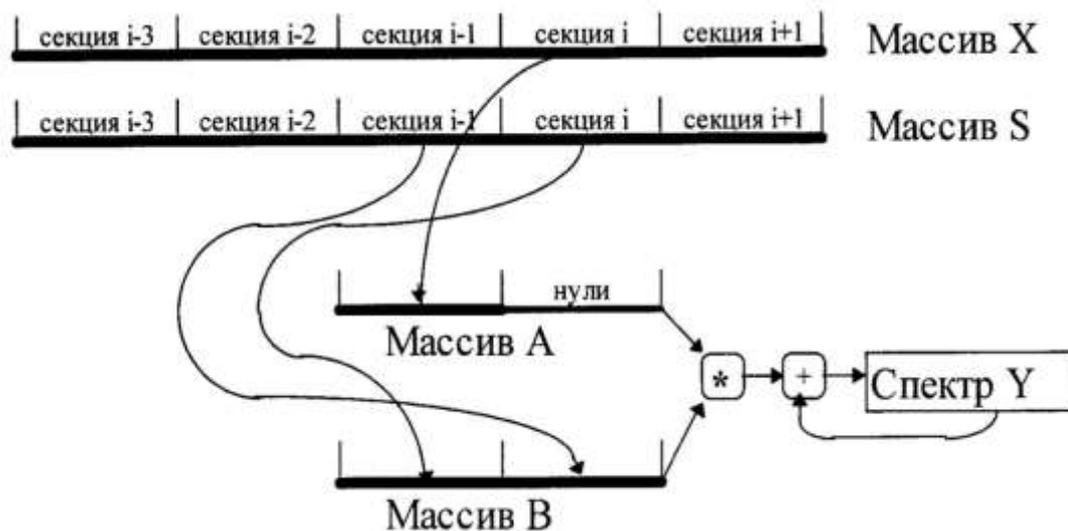


Рис.1

Вычислительные аспекты используемых методов.

Данная цифра включает только операции с плавающей точкой, и, следовательно, не учитывает вспомогательные операции, связанные с адресной арифметикой, пересылкой данных и т.д.

Трудоемкость вычисления свертки на основе БПФ без секционирования. Вычисление свертки в этом случае состоит из прямого БПФ, перемножения спектров и обратного БПФ. Размер Обрабатываемого массива равен $N+M$, где N – размер опорного, M – выходного массива. Общая трудоемкость:

$$T_{CB1} = 2T_{БПФ}(N + M) + 3 \cdot (N + M) = (N + M) \cdot (5 \cdot \log_2(N + M) - 8) + 16 \approx \\ \approx 5 \cdot (N + M) \cdot \log_2(N + M) \text{ флоп.}$$

Сюда не включается трудоемкость вычисления коэффициентов БПФ и спектра опорного сигнала, поскольку эти действия выполняются один раз за время работы программы, и при обработке большого числа входных файлов несут незначительную нагрузку.

Трудоемкость вычисления секционированной свертки на основе БПФ. В этом случае вычисление свертки состоит из вычислений прямого БПФ от каждой секции сигнала, умножения спектров на спектр опорного сигнала и прибавлением полученного произведения к текущему приближению спектра коррелограммы. По окончании обработки секций производится обратное БПФ от полученного спектра коррелограммы. Суммарная трудоемкость обработки по одному каналу:

$$T_{CB2} = N/M \cdot (T_{БПФ}(2 \cdot M) + 6 \cdot M) + T_{БПФ}(2 \cdot M) \approx \\ \approx N \cdot (5 \cdot \log_2 M + 2) + M \cdot (5 \cdot \log_2 M - 6) \approx 5 \cdot N \cdot \log_2 M \text{ флоп}$$

(подразумевается, что N делится на M нацело).

Трудоемкость вычисления коэффициентов БПФ и спектра опорного сигнала здесь также не учитывается.

Вывод: помимо всего прочего, секционирование позволяет еще и снизить суммарную трудоемкость вычисления свертки (однако следует помнить, что оно применимо только если $M \ll N$).

Объем памяти, необходимый для вычисления свертки без секционирования состоит из памяти для размещения обрабатываемого массива, опорного массива и массива коэффициентов БПФ и в сумме составляет $10 \cdot (N + M)$ байт. Здесь и далее подразумевается, что числа представляются в формате с плавающей точкой одинарной точности (формат float).

Объем памяти, необходимый для вычисления свертки с секционированием рассчитывается несколько по-другому, поскольку в этом случае за один цикл алгоритма обрабатываются данные по всем каналам. Поэтому используемая память из массива для обработки, массива для спектра одной секции опорного сигнала, вспомогательного массива для сохранения предыдущей секции опорного сигнала, массива коэффициентов БПФ и P массивов для накопления спектров коррелограмм, где P – число каналов. В сумме это составляет $22 \cdot M + 8 \cdot M \cdot P$ байт.

Точность оценивания ВКФ

Для ЛЧМ-сигнала вида

$$S(t) = a \cdot \cos \left(2\pi f_1 t + \frac{\gamma t^2}{2} \right) \quad 0 < t < T, \text{ где } \gamma = \frac{f_2 - f_1}{T}.$$

ВКФ имеет вид $y(t) = \int_0^T S(\lambda) \cdot S_{оп}(\lambda - t) \partial\lambda$. ВКФ с погрешностью

составляет $y^*(t) = \int_0^T S_{\Delta}(\lambda) \cdot S(\lambda - t) \partial\lambda$, где $S_{\Delta}(t) = S(f_2^*, f_1^*)$.

Здесь: $f_2^* = f_2 + \Delta f_2$, $f_1^* = f_1 + \Delta f_1$, $\Delta f_2 \ll f_2$, $\Delta f_1 \ll f_1$.

Введем погрешность вычисления ВКФ $\Delta y = y(t) - y^*(t)$. Для случая

$\Delta f_1 = 0,000$ и $\Delta f_2 = 0, 0.5, 1, 2, 5, 10$ процентов от заданного значения f_2 погрешность вычисления Δy представлена на рис. 2. Здесь $\Delta\gamma = (\Delta f_2 - \Delta f_1) T$

. Видна закономерность, что с ростом погрешности Δf_2 увеличивается длительность откликов и уменьшается их максимальная величина.

Заключение

1. В рамках сетевой технологии сбора и анализа данных в системах вибросейсмоакустического мониторинга предложены алгоритмы для обработки сигналов в темпе их поступления с применением секционирования входных и опорных сигналов. Оценены трудоемкости алгоритмов и объемы памяти для их исполнения.

2. Получена оценка величины ускорения по скорости исполнения программ для свертки в вариантах исполнения на универсальном процессоре – CPU и графическом – GPU.

3. Оценено влияние погрешности в формировании ЛЧМ-сигнала на качество ВКФ.

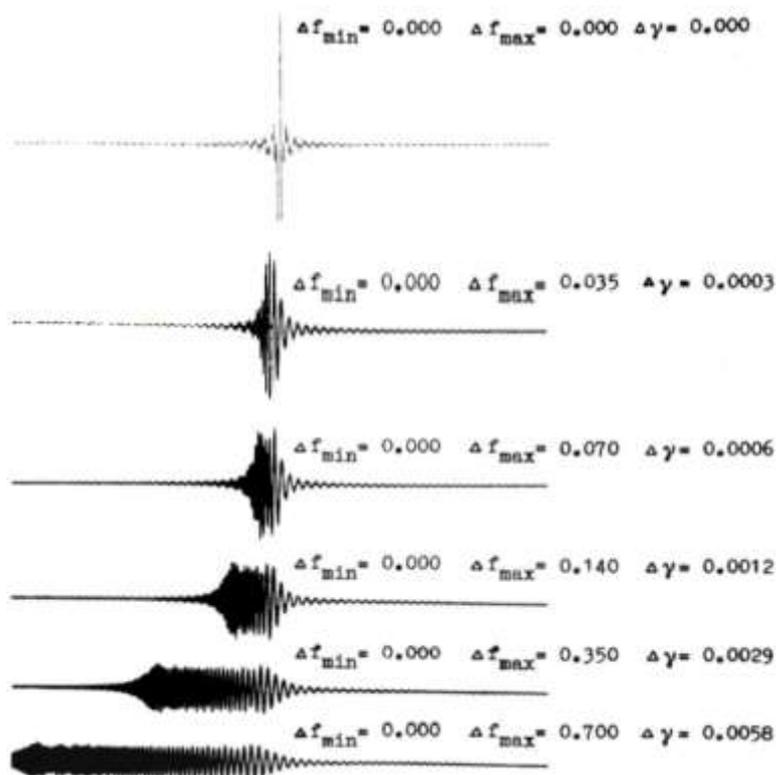


Рис. 2. Влияние погрешности в формировании ЛЧМ-сигнала на качество ВКФ.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Матвеев И.Н., Хайретдинов М.С. Организация сетевой архитектуры в задачах геофизического мониторинга. // Труды XI международной азиатской школы-семинара «Проблемы оптимизации сложных систем», 27 июля – 7 августа 2015 г. – г. Чолпон-Ата, 2015. – С.645–651.
2. Хайретдинов М.С., Геза Н.И., Ковалевский В.В., Седухина Г.Ф., Юшин В.И., Якименко А.А. Экспериментальная оценка абсолютных уровней когерентных сейсмических колебаний с помощью вибрационных технологий. // Технологии сейсморазведки. – Новосибирск, 2011. – №3. – С.84–92.
3. Губарев В.В., Ковалевский В.В., Авроров С.А., Воскобойникова Г.М., Седухина Г.Ф., Якименко А.А. Прогнозирование экологических рисков от взрывов по совокупности сопряженных геофизических полей. // Автометрия. – 2014. – Т. 50, № 4. – С. 3–13.

Спасибо за внимание!