сообщения
объединенного
института
ядерных
исследований
дубна

11-82-570

А.Кондор

ДИНАМИЧНАЯ СХЕМА ХРАНЕНИЯ БОЛЬШИХ РАЗРЕЖЕННЫХ МАТРИЦ ПРИ ОБРАБОТКЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

1982

1. ВВЕДЕНИЕ

При обработке данных эксперимента NA4 [1] пришлось столкнуться с проблемой создания схемы хранения больших /например, 300x300/ разряженных матриц*, которая в своей общей реализации может найти применение и в ряде других экспериментов. В частности, с учетом ограниченной разрешающей способности экспериментальной установки и ее влияния на определение структурных функций нуклонов некоторые искомые распределения рассеянных частиц являются решением следующей системы интегральных уравнений:

$$\int_{\Delta S_i} f(x)\,dx = \int_{\Delta S_i} dx \int_S u(x')\,p(x',x)\,dx' \qquad /1/$$
$$i = 1,2,\ldots,N,$$

где x' и x представляют собой какую-либо физическую величину; при этом x - измеренное значение, а штрихом обозначено идеальное /истинное/ значение; $u(x')$ - искомое распределение /например, распределение событий - актов рассеяния/; $p(x',x)$ - вероятность получения в эксперименте смещенной величины x, соответствующей истинной величине x'; $f(x)$ - смещенное распределение, о котором известен лишь его интеграл по ΔS_i - результат измерения в измерительном интервале ΔS_i ; ΔS_i - i-тый интервал измерения /подобласть изменения x или x' /, $i = 1,2,\ldots,N$; S - полная область возможных значений физической величины x, при этом $\bigcup_{i=1}^{N} \Delta S_i = S$.

При известной функции $p(x',x)$ уравнение /1/ решается путем применения стандартных методов алгебраизации интегральных уравнений и путем решения полученной при этом системы линейных уравнений.

Если вероятность $p(x',x)$ неизвестна в своей аналитической форме, то на основе известной информации об изучаемых в эксперименте процессах строится процедура /в общем случае методом

---

*Разреженной называют матрицу, имеющую малый процент ненулевых элементов. При этом относительно местоположения нулей никаких априорных предположений нет: они могут быть расположены совершенно случайным образом.

Монте-Карло/, которая генерирует переходы $x' \to x$. В ряде случаев методом Монте-Карло невозможно определить функции $p(x',x)$ с удовлетворительной точностью; тогда приходится находить способ алгебраизации, при котором нет необходимости иметь $p(x',x)$ в аналитической форме. В работе[2] предлагается метод обработки данных эксперимента, характеризуемого уравнением типа /1/, когда применяется матрица, элементам которой соответствует в наших обозначениях следующее выражение:

$$a_{ij} = \frac{\int\limits_{\Delta S_i} dx \int\limits_{\Delta S_j} u(x')\, p(x',x)\, dx'}{\int\limits_{\Delta S_j} u(x')\, dx'} . \qquad /2/$$

В настоящей работе излагается общая схема, обеспечивающая компактное динамичное изображение матрицы типа /2/ в быстродействующей памяти ЭВМ для случая, когда информация о вероятности $p(x',x)$ получается путем постепенного сбора переходов $x' \to x$, полученных из расчетов по методу Монте-Карло.

2. ОПРЕДЕЛЕНИЕ МАТРИЧНЫХ ЭЛЕМЕНТОВ И СХЕМА ИХ ХРАНЕНИЯ

В качестве примера рассмотрим случай двух переменных, т.е. когда величина x представляет собой две переменные, скажем, x_1 и x_2. Метод Монте-Карло генерирует переходы этих переменных из "истинных" x'_1 и x'_2 в "измеренные" x_1 и x_2. После генерации по методу Монте-Карло N_0 переходов распределения* физических величин x'_1 и x'_2 до перехода и x_1 и x_2 после него связываются следующим выражением:

$$n^a_{\ell,m} = \sum_{j=1}^{j_{max}} \sum_{k=1}^{k_{max}} M_{jk\ell m} = \sum_{j=1}^{j_{max}} \sum_{k=1}^{k_{max}} n^b_{jk} \frac{M_{jk\ell m}}{n^b_{jk}} =$$

$$= \sum_{j=1}^{j_{max}} \sum_{k=1}^{k_{max}} n^b_{jk}\, q_{\ell m jk}, \qquad /3/$$

где

$$q_{\ell m jk} = \frac{M_{jk\ell m}}{n^b_{jk}};$$

*Здесь и в дальнейшем сами распределения и их интегралы по измерительным подобластям будем различать только в случае, если это необходимо для понимания текста.

j_{max} - число подобластей по x'_1; k_{max} - число подобластей по x'_2; $n^a_{\ell m}$ - число событий после генерации переходов в областях измерения ℓ по x_1 и m по x_2; n^b_{jk} - число событий в областях j по x'_1 и k по x'_2; $M_{jk\ell m}$ - число событий, перешедших из области jk в область ℓm; $q_{\ell m jk}$ - доля событий, характеризующая переход $jk \to \ell m$.

Из формулы /3/ видно, что достаточно обеспечить экономное динамичное хранение матрицы $M_{jk\ell m}$ в силу того, что она полностью определяет $q_{\ell m jk}$ при помощи вектора n^b_{jk}.

На основе вышеприведенного сформулируем задачу, а именно - найти экономный способ изображения матрицы $M_{jk\ell m}$, число ненулевых элементов которой динамично возрастает при рассмотрении четырехмерных переходов. Исходя из начальных условий:

$$M_{jk\ell m} = 0, \quad n^b_{jk} = 0, \quad n^a_{\ell m} = 0,$$

$$j = 1 \div j_{max}, \quad k = 1 \div k_{max}, \quad \ell = 1 \div \ell_{max}, \quad m = 1 \div m_{max},$$

проведем следующие операции для всех переходов ($\pi = 1 \div N_0$): при переходе $jk \longrightarrow \ell m$

$$M_{jk\ell m} = M_{jk\ell m} + 1,$$

$$n^b_{jk} = n^b_{jk} + 1,$$

$$n^a_{\ell m} = n^a_{\ell m} + 1,$$

а при переходе $jk \to \emptyset$ /т.е. для неизмеряемых случаев/ M и n^a не меняются, а $n^b_{jk} = n^b_{jk} + 1$.

Специфичность схемы выражается именно в ее динамичности - число ненулевых элементов и порядок их неизвестны, в этом случае стандартные способы хранения[3] неприменимы для достаточно больших матриц.

3. ДИНАМИЧНАЯ СХЕМА ДЛЯ ИЗОБРАЖЕНИЯ МАТРИЦЫ M

Представим себе сетку, число узлов в которой равно $\ell_{max} \times m_{max}$. В каждом узле пусть будет расположена последующая сетка, центром которой является выбранный узел. Двоичная сетка изображена на рис.1. Конструкция вторичной сетки обеспечивает однозначное соответствие между элементом $M_{jk\ell m}$ матрицы M и одним узлом вторичной сетки, расположенной в узле ℓm первичной.

При накоплении элементов матрицы M, для каждой пары индексов ℓm находится величина $R_{\ell m}$ - радиус смещения, который определяется следующим образом:

$$R_{\ell m} = \max_{jk}(|j-\ell|, |k-m|),$$

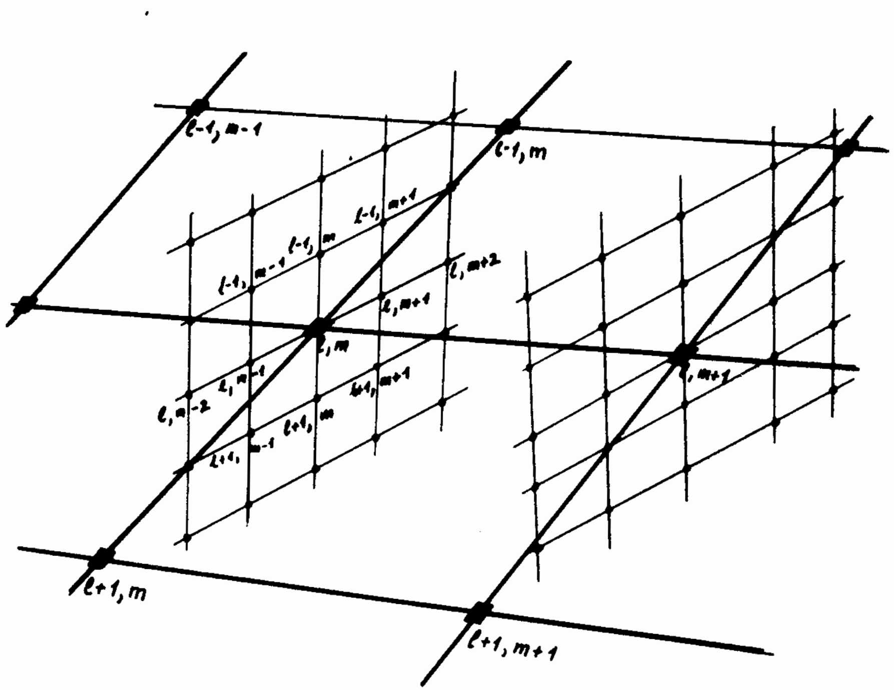

Рис.1. Двоичная сетка узлов для хранения матрицы M.

где j и k проходят по всем ненулевым элементам матрицы M с индексами ℓ и m. Таким образом, ненулевые элементы матрицы M можно поместить на числе мест

$$K = \sum_{\ell=1}^{\ell_{max}} \sum_{m=1}^{m_{max}} (2R_{\ell m} + 1)^2,$$

и для регистрации нахождения подматриц с индексами ℓm /т.е. вторичных сеток при первичном узле ℓm / достаточно иметь дальнейшие $\ell_{max} \times m_{max}$ мест.

На практике динамичная схема изображения четырехмерной матрицы реализуется следующим образом. Начинаем процедуру со значений $R_{\ell m} = 0$, $\ell = 1,2,\ldots,\ell_{max}$, $m = 1,2,\ldots,m_{max}$ и шаг за шагом увеличиваем $R_{\ell m}$ в зависимости от поступающих при генерации переходов $jk \to \ell m$. Подматрицы с радиусами смещения $R_{\ell m}$ скоро достигнут своих характерных размеров в зависимости от силы смещения, после чего наполнение элементов резко ускорится. На ЭВМ матрица M изображается одномерным массивом, а адреса подматриц хранятся отдельно. Величины $R_{\ell m}$ можно вычислить из адресов двух последовательных подматриц и с целью ускорения процедуры хранить отдельно.

4. ДИНАМИЧНАЯ СХЕМА ИЗОБРАЖЕНИЯ ЭКСТРЕМАЛЬНО БОЛЬШИХ МАТРИЦ M

В зависимости от эксперимента или выбранного числа экспериментальных подобластей некоторые $R_{\ell m}$ могут быть слишком большими для того , чтобы иметь возможность сохранять целую матрицу в оперативной памяти конкретной ЭВМ. В то же время переходы, соответствующие большим разностям $|j-\ell|$ или $|k-m|$, сравнительно редки. Это дает возможность решения проблемы с помощью внешней памяти. Представим себе, что в ходе сбора переходов один переход требует расширения подматрицы ℓ_{act}, m_{act} таким образом, что последнюю подматрицу ℓ_{last}, m_{last} придется вычеркнуть из оперативной памяти. Выпишем тогда эту подматрицу в отдельный файл /файл-1/, а переходы, которые при дальнейшем чтении входной информации будут относиться к ней, выпишем в другой файл /файл-2/. Таким образом, после прочтения полного входного файла переходов получаем три типа информации:

1. В оперативной памяти:
   часть матрицы $M_{jk\ell m}$, где $\ell = 1,2,\ldots,\ell^{(0)} \le \ell_{max}$
   $m = 1,2,\ldots,m^{(0)} \le m_{max}$.
2. В отдельном файле /файл-1/:
   часть матрицы $M_{jk\ell m}$, где $\ell = \ell^{(0)}+1,\ldots,\ell^{(1)} \le \ell_{max}$,
   $m = m^{(0)}+1,\ldots,m^{(1)} \le m_{max}$.
3. В отдельном файле /файл -2/:
   параметры переходов, соответствующие подматрицам, хранящимся в файле-1.

Дальнейшие шаги:

- часть матрицы M, находящейся в оперативной памяти, трансформируем в двухиндексную форму и выпишем в отдельный файл /файл-3/;
- введем в память часть M, сохраненную в файле-1, при этом $R_{\ell m} = 0$, $\ell = 1,2,\ldots,\ell^{(0)}$, $m = 1,2,\ldots,m^{(0)}$, поскольку для этих индексов все переходы уже были собраны в $M_{jk\ell m}$;
- прочитаем с файла-2 переходы и наполним $M_{jk\ell m}$;
- вторую часть матрицы $M_{jk\ell m}$ трансформируем также в форму двухиндексной матрицы и добавим ее к первой части $M_{jk\ell m}$ /матричные элементы $M_{jk\ell m}$- аддитивные величины/.

Из полной матрицы M и вектора n^b можно получить матрицу $q_{\ell m jk}$ или ее двухиндексный вариант q_{ts}. От многочисленных нулевых элементов q_{ts} можно освободиться следующим образом. Имея в виду, что $q_{ts} \ge 0$, любую строку t матрицы q:

a b c 0 d 0 0 0 e f 0 0 0 0 g...

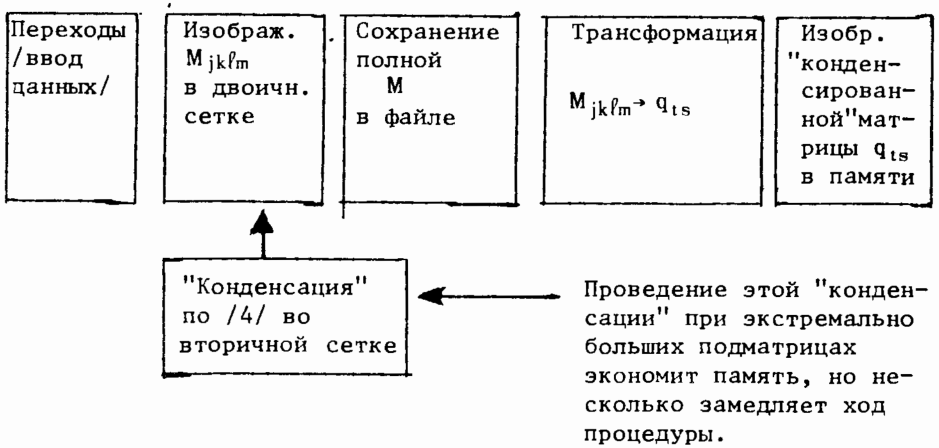

Рис.2. Блочная схема процедуры получения матриц M и q.

можно сохранить при уменьшенной длине в виде:

$$\overline{\text{a b c} -1 \text{ d } -3 \text{ e f } -4 \text{ g} \ldots} \qquad /4/$$

"Конденсация" типа /4/ обеспечивает значительную экономию, однако для дальнейшего применения пользователь должен выработать специальные процедуры матричных операций для таких конденсированных матриц.

Блочная схема полной процедуры изображена на рис.2.

5. РЕАЛИЗАЦИЯ НА ЭВМ СОХРАНЕНИЯ МАТРИЦ M и q

Процедура, предложенная в предыдущих параграфах, в настоящее время оформлена на языке фортран для ЭВМ CDC-6500.

Сама программа реализована для частного случая обработки данных эксперимента NA4 [4] - измерения дифференциальных сечений глубоконеупругого рассеяния мюонов при энергиях 120÷280 ГэВ и легко может быть приспособлена для обработки других экспериментов.

Сведения об этой программе будут опубликованы в отдельном сообщении для пользователей ЭВМ ОИЯИ.

В заключение хочется выразить благодарность участникам NA4 коллаборации за стимулирование проведения данных разработок.

ЛИТЕРАТУРА

1. Bollini D. et al. A Measurement of the Nucleon Structure Function from Muon-Carbon Deep Inelastic Scattering at High Q^2.Phys.Rev.Lett., 1981, 104B, p.403.
2. Kondor A. Method of Convergent Weights - a Possibility for Smearing Corrections, CERN-EP/NA4 Note 81-5.
3. Tewarson R.P. Sparse Matrices. Academic Press, 1973; Тьюарсон Р. Разреженные матрицы. "Мир", М., 1977.
4. Kondor A. QJKLM - a Code for the Raw Cross section Interaction Procedure. CERN-EP/NA4 Note 80-3.

Рукопись поступила в издательский отдел
21 июля 1982 года.

Кондор А. 11-82-570

Динамичная схема хранения больших разреженных матриц при обработке экспериментальных данных

Излагается схема хранения больших разреженных четырехмерных матриц, полученных при алгебраизации интегральных уравнений обработки данных физических экспериментов. Схема вырабатывалась в связи со специальными потребностями эксперимента №А4, но может найти широкое применение для любого эксперимента с заметными эффектами смещения. Программа, реализующая данную схему, обеспечивает удобство проведения матричных операций, а также вычисления физических характеристик установки, в том числе различные распределения и аксептанс.

Работа выполнена в Лаборатории ядерных проблем ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1982

Kondor A. 11-82-570

Dynamical Scheme for Storage of Large Sparse Matrices for Experimental Data Processing

A general scheme is presented for storage of fourfold-indexed sparse matrices appeared in discretization of integral equations of data processing procedure at physical experiments. The scheme is devoted to the peculiarities of the joint NA4 experiment (BCDMS collaboration CERN), but it can be implemented with minor efforts for any other experiment of large smearing effect. The program connected with this scheme gives possibilities to provide matrice operations or calculating physical characteristics of the experiment in particular distributions and acceptance as well.

The investigation has been performed at the Laboratory of Nuclear Problems, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1982

Перевод О.С.Виноградовой.