

На конкурс научных работ ОИЯИ за 2023 год

ПРЕДСТАВЛЕНИЕ

Представляется работа:

'Пакеты программ минимизации FUMILIM и накопления и обработки распределений NORA'

Раздел: **Научно-методические работы**

Коллектив соавторов:

1. Алексеев И.И.
2. Невский Д.В.
3. Селюгин О.В.
4. Ситник И.М

В представленный цикл работ входят 7 публикаций:

1. I.M. Sitnik, "Modification of the FUMILI Minimization Package preprint JINR E11-2008-43, Dubna
2. I.M. Sitnik, "Modification of the FUMILI Minimization Package-2 preprint JINR E11-2010-101, Dubna
3. I.M.Sitnik, "Development of the FUMILI Minimization Package Computer Physics Communications, 185(10) 2800 (2014)
4. I.M.Sitnik, "The new version of the FUMILIM Minimization Package Computer Physics Communications, v.209 p.199 (2016)
5. I.M.Sitnik, I.I.Alexeev, O.V.Selugin, "The final version of the FUMILIM Minimization Package Computer Physics Communications, v.251 p.107202 (2020)
6. I.M.Sitnik, I.I.Alexeev, D.V.Nevsky, "Debugging the FUMILIM minimization package" Computer Physics Communications, v.294, p.108886 (Jan 2024)

7. I.M. Sitnik, "5D Histogram package NORA"
Computer Physics Communications, v.283 p.108560 (2023)

Представляемый цикл работ выполнен в рамках тем 02-0-0941-91/2009, 02-1-1097-2010/2012, 02-1-1097-2010/2021 проблемно-тематического плана ОИЯИ.

Пакет программ FUMILIM является данью памяти выдающемуся математику, профессору **Игорю Николаевичу Силину**, создавшему первый в мире пакет минимизации FUMILI [1]. Пакет программ NORA создавался в течение многих лет, начиная с 1978 года [2].

При развитии пакета FUMILI ставилась цель расширить класс решаемых задач, облегчить интерфейс пользователя и повысить скорость решения задач. Пакет NORA также расширяет класс решаемых задач за счет повышения допустимых размерностей распределений, а также увеличивает число возможных операций для объектов всех размерностей по сравнению с известными пакетами ROOT и HBOOK.

1 Пакет FUMILIM

На премию по пакету FUMILIM выдвигаются:

1. Возможность работы с мульти-наборными задачами с неограниченным числом наборов.

Все чаще в современном мире встречаются огромные множества экспериментальных точек, разделенные на подмножества (такое может быть и с первичными данными ([3, 4]) и с конечными ([5])). Каждое подмножество описывается функцией, зависящей от глобальных параметров (общие для всего множества) и локальных (для каждого подмножества свой набор).

Таковыми задачами оказались настройка оффсетов дрейфовых трубок[3] (144 параметра, 325 подмножеств, 137000 эксп. точек, выдача дана в файле "stroffset.txt"), Задача со спутниками на орбите Меркурия [4] содержала до 1000 параметров.

Полное число параметров в такой задаче

$$N_t = N_g + N_l N_s$$

где N_g – число глобальных параметров, N_l – число локальных параметров, N_s – число подгрупп (витков), которое не имеет ограничения сверху. В каждой экспериментальной точке ищутся только

$$N_{gt} = N_g + N_l.$$

параметров. Каких именно, выясняется на этапе самонастройки программы фита. В задаче МГУ $N_g = 6$, $N_l = 7$.

При $N_{gt} > 25$, как и для всех пакетов минимизации, нарастают проблемы сходимости.

Задача [5] имеет дело с наборами данных, полученных в разных экспериментах. Имеется 18 общих параметров и 90 локальных (нормировочный коэффициент для набора).

2. Возможность работы с многомерными экспериментальными точками, описываемыми векторными функциями. Здесь мы приведем в качестве примера опять же задачу МГУ [4], где аргументом является время, а значениями - три компоненты магнитного поля.
3. Автоматический выброс плохих точек.

В больших массивах экспериментальных точек неизбежно присутствуют "плохие" точки (аномально большой вклад в χ^2), которые могут серьезно исказить результаты фита.

На рисунках показан выброс "плохих" точек при измерении поляризации пучка на нуклотроне и при фите магнитного поля "Меркурия" на одном из витков. Выброс плохих точек вручную в первом случае – посильная задача, во втором – нереальная (4000 витков, 1000 эксп. точек в каждом витке).

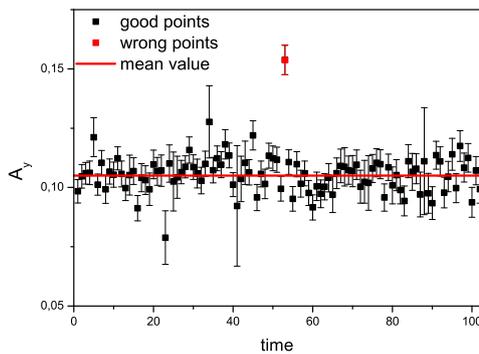


Рис. 1: Measurement of beam polarization.

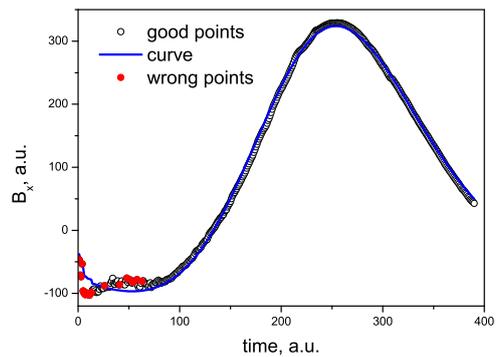


Рис. 2: Фит магнитного поля во время одного из оборотов спутника.

4. Реконструкция событий.

Для обеспечения скорости счета в multi-call задачах созданы две сверхбыстрые подпрограммы (для 2-х и для произвольного числа параметров).

При проведении треков через однозначные детекторы достигается скорость 10^6 трек/сек при частоте процессора 2.5 Мггц.

Для дрейфовых детекторов скорость составляет $6 \cdot 10^5$ трек/сек. Треки проводились через трубки в [3] и через дрейфовые камеры при обработке данных по проектам ALPOM2, STRELA.

5. Параллельный фит для тяжелых функций пользователя.

Распараллеливается взятие производных по параметру. Таким образом, оптимальное число ядер процессора равно числу свободных параметров для подмножества, для которых не написаны аналитические производные. Оптимальное число ядер процессора для задачи [5] – $\simeq 20$. При числе ядер 4, фактор ускорения достиг 2.5.

6. Поиск глобального минимума.

Многие задачи имеют значительное число локальных минимумов для функционала $\chi^2 = F(a_i, \dots)$, где a_i – параметры. В этом случае, при разных начальных значениях параметров могут быть получены разные конечные значения, так как стандартная процедура фита уходит в ближайший локальный минимум. Для таких случаев предусмотрен прямой перебор значений выбранных параметров в заданных пределах с заданным шагом. Найденные таким образом значения параметров в силу конечного значения шага являются приближительными, так что по завершении этой процедуры программа переходит к стандартному фиту. Для этого весьма времязатратного процесса найдены способы оптимизации, в том числе, распараллеливание. Эта процедура использовалась в задаче [5].

7. Широкий набор встроенных функций для комфортной работы с полученными в эксперименте данными.

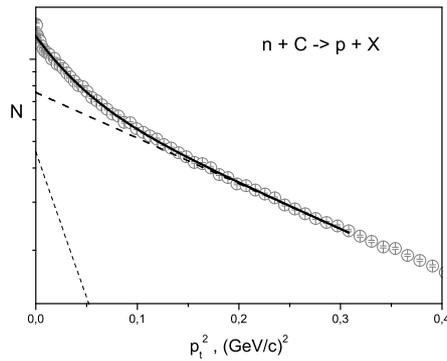


Рис. 3: nC scattering distribution. Black line – sum of 2 exponents. Dashed lines – components.

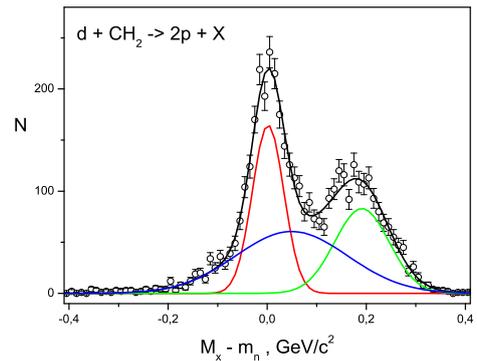


Рис. 4: 3 Gauss for two-heads peak. Black line – sum of 3 Gauss. Color lines – components.

На рисунках 3,4 показаны экспериментальные результаты по проектам ALPOM2 [6], STRELA[7]. В каждом случае набор кривых на рисунке обеспечен одной строкой обращения к пакету со ссылкой на массив экспериментальных данных. **Задание начальных значений параметров не требуется.** В эксперименте по рассеянию нейтрона на разных мишенях (ALPOM2) обнаружено 2 экспоненциальных наклона в реакции перезарядки $n \rightarrow p$.

8. Возможность в функции пользователя определять аналитически произвольное число производных по параметрам.

В исходном пакете FUMIL (а также MINUIT) можно определить или все производные, или ни одной. Между тем, аналитическое определение любого числа производных всегда желательно, так как это повышает одновременно надежность и скорость фита. Правильное написание производных контролируется при самонастройке программы фита с помощью численного взятия производной. При серьезном расхождении выдается предупреждение.

9. Создан дружественный интерфейс пользователя, позволяющий легко настраиваться на разные форматы входных данных и управлять выходными данными, менять режимы фита. Значительное внимание уделено защите от ошибок пользователя. Каждая ошибка сопровождается строкой диагностики.

Пакеты FUMIL и MINUIT имеет ограничение на число параметров в функции пользователя (<101). Приведенные выше ссылки на задачи на несколько сот параметров, показывают, что **существует класс задач, для решения которых пакету FUMILIM нет альтернативы.**

Пакет FUMILIM использовался в премиальной работе [6] и лежит в основе премиальной работы [4].

Подробно с пакетом можно ознакомиться в User Guide (прилагается).

2 Пакет NORA

На премию по пакету NORA выдвигаются:

1. Повышение размерности распределений до 5 (против 3 у ROOT).

4D распределения вторичного пучка накапливались на установке ALPOM2 (x , dx/dz , y , dy/dz) для последующей симуляции методом Монте-Карло для определения геометрической эффективности установки. 4D объект не раскладывается на $2*2D$ из-за высокой корреляции между x - и y -компонентами пучка.

5D объекты были апробированы в эксперименте [6] на установке ALPOM2. Гистограммировались следующие величины: $\cos \varphi$, $f(\theta)$, сигнал калориметра, мода поляризации, пронумерованные импульсы пучка. В программе следующего этапа 5D объект расщеплялся на 30 2D объектов. Время расщепления объектов составляет доли секунды. Время накопления распределений кратно числу распределений. В то же время, работы 5D- и 2D-накопителей отличается всего на 15%. Таким образом, выигрыш времени на этапе накопления распределений составляет $30/1.15$, который не декомпенсируется потерей времени на следующем этапе обработки.

2. Конволюция распределений.

Для корректного вычитания спектров рассеяния на разных мишенях нужно эти спектры выровнять по показателю многократного рассеяния (σ). Конволюция проводится для спектра наименьшим показателем многократного рассеяния (σ_1) с параметром

$$\sigma = \sqrt{\sigma_2^2 - \sigma_1^2}.$$

На рис.5 показаны спектры на CH_2 - и пустой (оригинальный и преобразованный) мишенях, добытых на установке ALPOM2. На рис.6 показано вычитание спектров. Видно, что прямое вычитание дает нефизический результат в виде отрицательных значений вычитания. В то же время видно, что при вычитании модифицированного спектра область значений аргумента, которой можно доверять, расширяется в сторону меньших значений аргумента.

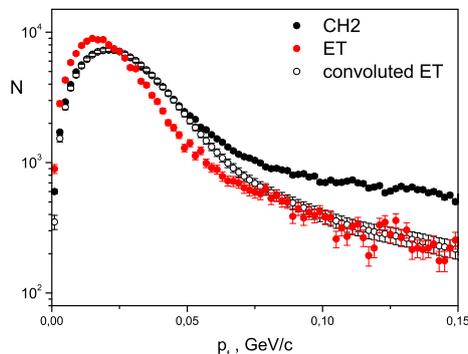


Рис. 5: Source and convoluted spectra.

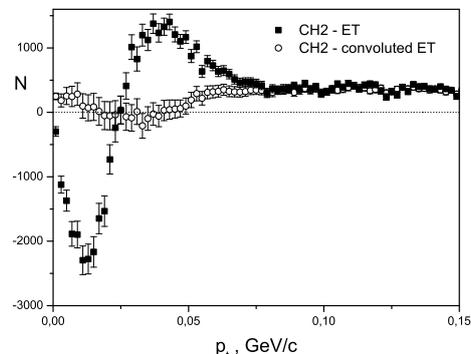


Рис. 6: Subtracting an empty target.

3. x-преобразования.

В программе накопления распределений (ALPOM2) использовалась переменная $p_t = p_0 \sin \theta$. Спектры показаны на рис.7. Фит этих спектров показал хорошее описание с помощью суммы двух рэлеевских распределений. Это навело на мысль, что более иллюстративно будет показывать спектры в зависимости от другой переменной (Мандельстама) $t \simeq -p_t^2$. Один из преобразованных спектров уже показан на рис.3, где четко различаются два наклона. Фит этого распределения при помощи суммы двух экспонент дает такие же наклоны, как и предыдущий фит.

Таким образом, наличие x-преобразований позволяет избежать возврата к стадии накопления распределений, если первоначально переменные выбраны не слишком удачно.

Предварительная версия пакета NORA использовалась в премиальной работе [6]. Подробно с пакетом можно ознакомиться в User Guide (прилагается).

3 Статистика использования пакетов

На 9 окт.2023 г общее число цитирований всех версий FUMILIM - 32 (данные ScienceDirect), последняя версия (CPC, Jan 2024) уже имеет 1046 загрузок (<https://data.mendeley.com/datasets/xphc8ctxr3/3>)

Пакет NORA имеет 83 загрузки (<https://data.mendeley.com/datasets/k363z3kp3x/1>).

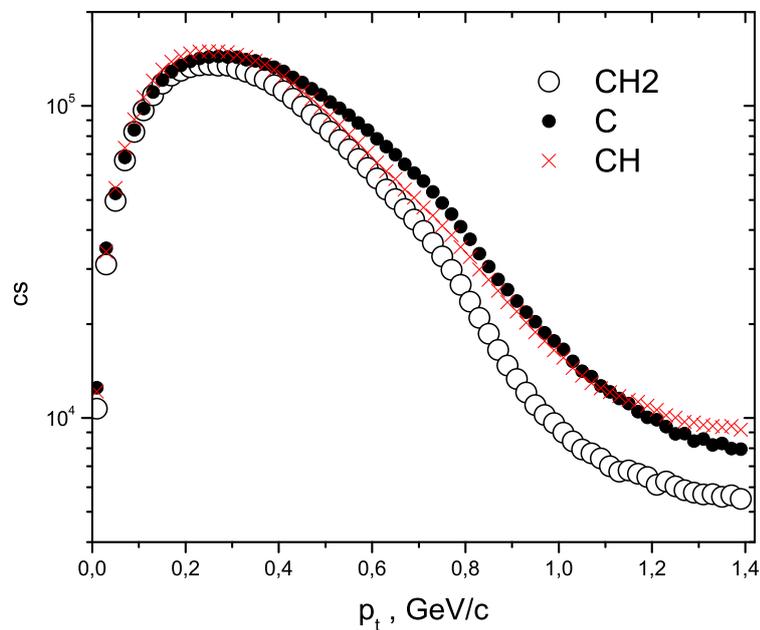


Рис. 7: Спектры рассеяния нейтронов на разных мишенях.

Список литературы

- [1] Silin I.N. CERN Program Library, D510, FUMILI
- [2] G.L. Maznyu, I.M. Sitnik, E.A. Strokovski,
JINR, D10,11-11264, Dubna, 1978 (in Russian).
- [3] I.M. Sitnik et al., Phys.Rev. C 84 034006 (2011).
- [4] I.I. Alexeev, et al., Icarus 209 (2010) 23.
Работа отмечена премией Ломоносова за 2011 год.
- [5] O.V. Selugin, will be submitted soon.
- [6] S.N. Basylev et al., Eur. Phys. J. **A56** (2020) 26.
Работа отмечена 1-й премией ОИЯИ за 2020 год.
- [7] S.N. Basylev et al., Eur. Phys. J. **A57** (2021) 133.

Председатель НТС ЛФВЭ

Е.А. Строковский

Ученый секретарь НТС ЛФВЭ

С.П. Мерц