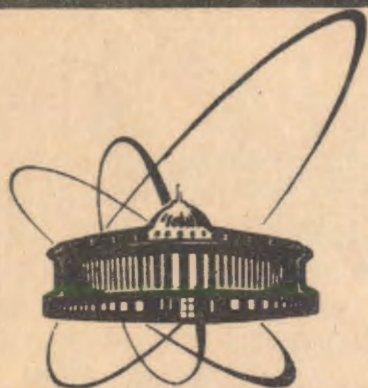


91-483



ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА

P10-91-483

А. Ж. Кетикян*, Е. В. Комиссаров, В. С. Курбатов,
И. Н. Силин

КИНЕМАТИЧЕСКИЙ ФИТ: НЕТРАДИЦИОННЫЙ
ПОДХОД К ТРАДИЦИОННОЙ ПРОБЛЕМЕ

Направлено в журнал "Nuclear Instruments
and Methods"

*Ереванский физический институт

1991

1. Введение

Тридцать лет тому назад, в эпоху расцвета методики пузырьковых камер, были опубликованы работы [1,2], на основе которых в практику эксперимента был введен кинематический фит. Впоследствии было создано немало систем обработок данных, которые структурно были похожи друг на друга и выглядели примерно так: вначале проводилась так называемая геометрическая реконструкция событий, т.е., имея координаты треков, находили углы, импульсы, координаты вершин событий и т.д. Затем (и это была функция кинематического фита), делая определённые предположения о типе процесса, проводили его идентификацию с привлечением законов сохранения. Технически это выглядело так: записывался χ^2 -функционал, в котором фигурировали в явном виде оценки параметров, найденных в процессе геометрической реконструкции, и их матрица ошибок, и затем проводилась его минимизация с наложением уравнений связи, выражавших законы сохранения. Минимизация проводилась различными способами: метод множителей Лагранжа [1,2] или метод с использованием дополнительного квадратичного члена [3].

Наряду с таким подходом к системе обработки данных, возможен и другой подход: на первой стадии проводить распознавание события, т.е. ассоциацию измеренных координат в событие, а на второй проводить кинематическую идентификацию события, используя непосредственно измеренные координаты. На наш взгляд, второй подход иногда предпочтительнее в связи с меньшим риском искажения информации. Кроме того, такая постановка задачи представляет самостоятельный интерес для общего случая минимизации функционала со связями.

При использовании методики минимизации с применением множителей Лагранжа в традиционном подходе функционал записывался в виде:

$$M = \sum_{i,j} (x_i - x_i^a) \cdot S_{ij} \cdot (x_j - x_j^a) + 2 \cdot \sum_{\lambda} \alpha_{\lambda} \cdot f_{\lambda}(X), \quad (1)$$

где x_i^a - найденные на стадии реконструкции значения параметров: углов, импульсов и т.д. S_{ij}^{-1} - их весовая

матрица, X_i - "истинные", т.е. искомые значения тех же параметров, а $f_\lambda(\vec{X})$ - уравнения связи.

В предлагаемом подходе функционал имеет вид

$$M = \sum_{i,j} [C_i(\vec{X}) - C_i^m] \cdot G_{i,j}^{-1} \cdot [C_j(\vec{X}) - C_j^m] + 2 \cdot \sum_{\lambda} \alpha_{\lambda} \cdot f_{\lambda}(\vec{X}). \quad (2)$$

Здесь C_i^m - найденные на стадии распознавания координаты, $G_{i,j}^{-1}$ - их ковариационная матрица, \vec{X} - вектор искомых параметров (углы, импульсы, координаты вершин...), $C_i(\vec{X})$ - "истинные" значения координат, которые являются известными функциями параметров.

Целью настоящей работы является изложение методики минимизации функционалов типа (2), даются необходимые формулы, матрицы ошибок и т.д.

Даются также результаты использования этой методики.

2. Метод

2.1. Линейный случай

Рассмотрим вначале линейный случай, т.е. когда функции $C_i(\vec{X})$ и $f_{\lambda}(\vec{X})$ линейны:

$$C_i(\vec{X}) = C_i^0 + (DC \cdot \vec{X})_i \longrightarrow C(\vec{X}) = C^0 + DC \cdot \vec{X}; \quad i = 1 + nf, \quad (3)$$

$$f_{\lambda}(\vec{X}) = F_{\lambda}^0 + (DF \cdot \vec{X})_{\lambda} \longrightarrow f(\vec{X}) = F^0 + DF \cdot \vec{X}; \quad \lambda = 1 + nc. \quad (4)$$

Здесь \vec{X} - вектор искомых параметров, число компонент np ,

DC - прямоугольная матрица, у которой nf строк и np столбцов,

DF - прямоугольная матрица, размером $nc \times np$.

Преобразуем функционал (2), используя формулы (3):

$$\begin{aligned}
c_i(\vec{X}) - c_i^m &= (c_i^0 - c_i^m) + \sum_k DC_{ik} \cdot X_k, \\
\sum_{i,j} [c_i(\vec{X}) - c_i^m] \cdot G_{ij} \cdot [c_j(\vec{X}) - c_j^m] &= \\
= \sum_{i,j} [(c_i^0 - c_i^m) + \sum_k DC_{ik} \cdot X_k] \cdot G_{ij} \cdot [(c_j^0 - c_j^m) + \sum_l DC_{jl} \cdot X_l] &= \\
= T_0 + 2 \cdot \sum_k T_k \cdot X_k + \sum_{k,l} X_k \cdot H_{kl} \cdot X_l. & \quad (5)
\end{aligned}$$

Здесь

$$T_0 = \sum_{i,j} (c_i^0 - c_i^m) \cdot G_{ij} \cdot (c_j^0 - c_j^m),$$

$$T_k = \sum_{i,j} (c_i^0 - c_i^m) \cdot G_{ij} \cdot DC_{jk},$$

$$H_{kl} = \sum_{i,j} DC_{ki}^T \cdot G_{ij} \cdot DC_{jl}.$$

В общем виде функционал M :

$$M = T_0 + 2 \sum_k T_k \cdot X_k + \sum_{k,l} X_k \cdot H_{kl} \cdot X_l + 2 \sum_{\lambda} \alpha_{\lambda} \cdot f_{\lambda}(\vec{X}). \quad (6)$$

Найдём искомые значения \vec{X}^e :

$$\frac{\partial M}{\partial X_k} = 2 \cdot [T_k + \sum_l H_{kl} \cdot X_l + \sum_{\lambda} DF_{k\lambda}^T \cdot \alpha_{\lambda}] = 0. \quad (7.1)$$

$$\frac{\partial M}{\partial \alpha_{\lambda}} = 2 \cdot [F_{\lambda}^0 + (DF \cdot \vec{X})_{\lambda}] = 0. \quad (7.2)$$

Верхний индекс T означает транспонирование матрицы.

Запишем уравнения (7) в матричном виде:

$$\vec{T} + H \cdot \vec{X} + DF^T \cdot \vec{\alpha} = 0, \quad (8.1)$$

$$F^0 + DF \cdot \vec{X} = 0. \quad (8.2)$$

Выразим \vec{X} через $\vec{\alpha}$, используя (8.1):

$$\vec{X} = -H^{-1} \cdot \vec{T} - H^{-1} \cdot DF^T \cdot \vec{\alpha} \quad (9)$$

Подставим (9) в (8.2):

$$\vec{F}^0 - DF \cdot H^{-1} \cdot \vec{T} - DF \cdot H^{-1} \cdot DF^T \cdot \vec{\alpha} = 0$$

или

$$\vec{\alpha} = R^{-1} \cdot \vec{F}^0 - R^{-1} \cdot DF \cdot H^{-1} \cdot \vec{T} \quad (10)$$

Наконец,

$$\vec{X}^e = -H^{-1} \cdot DF^T \cdot R^{-1} \cdot \vec{F}^0 + [H^{-1} \cdot DF^T \cdot R^{-1} \cdot DF - 1] \cdot H^{-1} \cdot \vec{T},$$

$$R = DF \cdot H^{-1} \cdot DF^T. \quad (11)$$

Найдём матрицы ошибок. Используя соотношения

$$E(C_i^m) = C_i^0 + (DC \cdot \vec{X}^t)_i, \quad (12)$$

$$E\left\{ [C_i^m - E(C_i^m)] [C_j^m - E(C_j^m)] \right\} = G_{ij}^{-1}$$

и определение \vec{T} (см. формулы (5)), нетрудно показать, что

$$\vec{X}^e = \vec{X}^t + [H^{-1} \cdot DF^T \cdot R^{-1} \cdot DF - 1] \cdot H^{-1} \cdot DC^T \cdot G \cdot [E(C^m) - C^m],$$

$$\vec{\alpha}^e = -R^{-1} \cdot DF \cdot H^{-1} \cdot DC^T \cdot G \cdot [E(C^m) - C^m]. \quad (13)$$

Здесь \vec{X}^t - истинные значения вектора параметров \vec{X} .

Формулы (13) полезны для получения матриц ошибок. С их помощью получим:

$$E(\vec{X}^e) = \vec{X}^t,$$

$$E(\vec{\alpha}^e) = 0. \quad (14)$$

т.е. оценки параметров \vec{X}^e - несмещённые, соответственно,

$$\langle (X_i^e - X_i^t)(X_j^e - X_j^t) \rangle = (H^{-1} - H^{-1} \cdot DF^T \cdot R^{-1} \cdot DF \cdot H^{-1})_{ij},$$

$$\langle \alpha_i^e \cdot \alpha_j^e \rangle = R_{ij}^{-1}. \quad (15)$$

$$\langle (X_i^e - X_i^t) \cdot \alpha_j^e \rangle = 0.$$

Наконец, найдём $E(\chi^2(\vec{X}^e))$:

$$\chi^2 = T_0 + 2 \cdot \vec{T} \cdot \vec{X}^e + (\vec{X}^e)^T \cdot H \cdot \vec{X}^e, \quad (16)$$

$$E(\chi^2) = n f - n p + n c. \quad (17)$$

В традиционном подходе, когда $n f = n p$, формула (17) даёт

$$E(\chi^2) = n c.$$

2.2. Общий случай

Рассмотренная выше линейная модель практически никогда не встречается. В общем случае функции $S(X)$ и $r(X)$ существенно нелинейны. В линейной модели решение уравнений (7) находится сразу, а в общем случае - посредством итераций. Запишем уравнения, которым должно удовлетворять решение (в рамках метода множителей Лагранжа):

$$\frac{\partial M}{\partial X_k} = 2 \cdot \sum_{i,j} \frac{\partial C_i}{\partial X_k} \cdot G_{ij} \cdot [C_j(X) - C_j^m] + 2 \cdot \sum_{\lambda} \alpha_{\lambda} \frac{\partial r_{\lambda}}{\partial X_k} = 0, \quad (18.1)$$

$$\frac{\partial M}{\partial \alpha_{\lambda}} = 2 \cdot r_{\lambda}(X) = 0. \quad (18.2)$$

Предположим, что нам известны параметры на v -й итерации \vec{X}^v . Рассмотрим значение функции $\frac{\partial M}{\partial X_k}$ в точках $X = \vec{X}^{v+1}$ и

$\vec{a} = \vec{a}^{\nu+1}$ вблизи значений $\vec{x} = \vec{x}^{\nu}$.

$$\frac{\partial M(\vec{x}^{\nu+1}, \vec{a}^{\nu+1})}{\partial x_k} \cong 2 \cdot \sum_{i,j} \frac{\partial c_i}{\partial x_k} \cdot G_{ij} \cdot [c_j(\vec{x}^{\nu}) - c_j^m] +$$

$$2 \cdot \sum_{\lambda} \alpha_{\lambda}^{\nu+1} \cdot \frac{\partial f_{\lambda}}{\partial x_k} + 2 \cdot \sum_{i,j,l} \frac{\partial^2 c_i}{\partial x_k \cdot \partial x_l} \cdot G_{ij} \cdot [c_j(\vec{x}^{\nu}) - c_j^m] \cdot \delta x_l +$$

$$2 \cdot \sum_{i,j,l} \frac{\partial c_i}{\partial x_k} \cdot G_{ij} \cdot \frac{\partial c_j}{\partial x_l} \cdot \delta x_l + 2 \cdot \sum_{\lambda} \alpha_{\lambda}^{\nu+1} \cdot \frac{\partial^2 f_{\lambda}}{\partial x_k \cdot \partial x_l} \cdot \delta x_l, \quad (19)$$

где $\delta x_l = x_l^{\nu+1} - x_l^{\nu}$.

Все производные в формуле (19) рассчитываются в точке $\vec{x} = \vec{x}^{\nu}$.

В соответствии с методом линеаризации [4] нам следует пренебречь вторыми производными (естественно, если нет вырождения по первым производным (см. работу [4])) и вместо (18) следует записать

$$\frac{\partial M}{\partial x_k} \cong \sum_{i,j} \frac{\partial c_i}{\partial x_k} \cdot G_{ij} \cdot [c_j(\vec{x}^{\nu}) - c_j^m] +$$

$$2 \cdot \sum_{i,j,l} \frac{\partial c_i}{\partial x_k} \cdot G_{ij} \cdot \frac{\partial c_j}{\partial x_l} \cdot \delta x_l + \sum_{\lambda} \alpha_{\lambda}^{\nu+1} \cdot \frac{\partial f_{\lambda}}{\partial x_k} = 0, \quad (20)$$

$$f_{\lambda}(\vec{x}^{\nu}) + \sum_l \frac{\partial f_{\lambda}}{\partial x_l} \cdot \delta x_l = 0.$$

Необходимо прокомментировать факт пренебрежения вторыми производными. Можно ожидать, что вклад от третьего члена в правой части (19) вблизи искомого решения мал за счёт малости и знакопеременности входящих в сумму членов $c_j(\vec{x}^{\nu}) - c_j^m$. Однако про последний член в (19) заранее сказать что-либо определённое трудно (не известны величины α_{λ}). Тем не менее малость вклада от последнего члена можно предположить по косвенным соображениям: в принципе уравнения связи можно

учитывать методом исключения m - s параметров, выражая их через остальные n - s параметры. Метод исключения параметров математически эквивалентен методу множителей Лагранжа с точки зрения конечного результата и при его использовании подобные вопросы не возникают.

Сравнивая уравнения (20) с соответствующими уравнениями линейной модели, мы увидим их одинаковость, если положим

$$c_i^0 = c_i(\vec{x}^{\nu}) - \sum_l \frac{\partial c_i(\vec{x}^{\nu})}{\partial x_l} \cdot x_l^{\nu}, \quad (21)$$

$$F_i^0 = f_i(\vec{x}^{\nu}) - \sum_l \frac{\partial f_i(\vec{x}^{\nu})}{\partial x_l} \cdot x_l^{\nu} = (f - DF \cdot \vec{x}^{\nu})_i.$$

Соответственно, справедливы и все остальные формулы, которые были получены для линейной модели.

Нетрудно получить и итерационные формулы:

$$\vec{x}^{\nu+1} = \vec{x}^{\nu} - H^{-1} \cdot DF^T \cdot R^{-1} \cdot \vec{f}^{\nu} +$$

$$+ [H^{-1} \cdot DF^T \cdot R^{-1} \cdot DF - 1] \cdot H^{-1} \cdot DC^T \cdot G \cdot [\vec{c}^{\nu} - \vec{c}^m], \quad (22)$$

$$\vec{a}^{\nu+1} = R^{-1} \cdot \vec{f}^{\nu} - R^{-1} \cdot DF \cdot H^{-1} \cdot DC^T \cdot G \cdot [\vec{c}^{\nu} - \vec{c}^m].$$

Здесь $\vec{f}^{\nu} = \vec{f}(\vec{x}^{\nu})$, $\vec{c}^{\nu} = \vec{c}(\vec{x}^{\nu})$.

При получении этих формул полезно использовать следующее представление векторов \vec{F}^0 и \vec{T} :

$$\vec{F}^{\nu} = \vec{f}^{\nu} - DF \cdot \vec{x}^{\nu},$$

$$\vec{T}^{\nu} = DC^T \cdot G \cdot [\vec{c}^{\nu} - \vec{c}^m] - H \cdot \vec{x}^{\nu}.$$

3. Контроль сходимости

К сожалению, в методе множителей Лагранжа невозможен контроль последовательного улучшения приближений, применяемый в обычных методах оптимизации - по улучшению значений оптимизируемой функции (функционал (1) имеет в окрестности решения седловую форму). Поэтому, вообще говоря, нужно независимо контролировать качество описания экспериментальных данных и точность удовлетворения уравнений связи. Будем считать итерацию успешной, если в результате её выполнения уменьшается либо χ^2 по экспериментальным точкам, либо сумма квадратов взвешенных¹ невязок уравнений связи. Если ни одно из этих условий не выполняется, шаг следует укоротить, сохраняя направление движения. Требуется, чтобы невязки всегда уменьшались, нецелесообразно, т.к. мы можем начать двигаться слишком мелкими шагами из-за нелинейности связей. χ^2 в направлении шага может увеличиваться, т.к. этого может потребовать удовлетворение связей. Так же, как и в^[4], целесообразна параболическая интерполяция во избежание незатухающих колебаний вокруг решения. Полезно укорачивать шаг даже и в том случае, когда уменьшающаяся сумма уменьшается мало (например, меньше чем на половину глубины аппроксимирующей параболы). Может потребоваться интерполирование обеих контролируемых сумм, в этом случае следует выбирать вариант, дающий в результате интерполяции большие шаги.

Развитый выше формализм был проверен на экспериментальном материале, полученном на установке ИСТРА-М^[5] в апрельском сеансе 1990 года. Для отработки

¹ Взвешивание связей необходимо для выравнивания их масштабов. Естественными весами являются веса, обратные оценкам формально вычисленных ошибок уравнений связи. Из-за нелинейности уравнений связи веса нужно пересчитывать в начале каждой итерации, но сравнивать начальные и конечные значения невязок нужно при одинаковых весах.

методики был выбран тестовый распад $K_{\pi 2}^- \rightarrow \pi^- \pi^0$, в котором регистрировался пучковый K^- -мезон (либо только направление, либо направление и абсолютная величина импульса), направление вторичного π^- -мезона (абсолютная величина импульса не измерялась) и направления и энергии γ -квантов от распада π^0 -мезона. Это означало, что в зависимости от того, была или не была измерена абсолютная величина импульса K^- -мезона, мы имеем 4с- или 3с-фит. В качестве начального приближения выбирались значения параметров, полученные на стадии распознавания события.

Рассмотренный выше механизм контроля сходимости был реализован не полностью, а именно: на каждой итерации вычислялась матрица R (формула (11)), для которой справедливо следующее соотношение:

$$\langle \delta f_i, \delta f_j \rangle = R_{ij},$$

где f_i - уравнение связи с номером i .

Соответственно, $\frac{1}{\langle \delta f_i^2 \rangle} = R_{ii}^{-1}$. Рассчитывалась невязка уравнений связи:

$$vrf0 = \sum_i \frac{f_i^2}{\langle \delta f_i^2 \rangle} = \sum_i f_i^2 \cdot R_{ii}^{-1} \quad (23)$$

при значениях параметров \vec{x}^p . Та же невязка рассчитывалась при новых значениях параметров \vec{x}^{p+1} (формула (21)) - обозначим её $vrf1$. Значение элементов R_{ii}^{-1} было, естественно, одним и тем же при расчёте $vrf0$ и $vrf1$. Если $vrf1 > vrf0$, то приращения уменьшались вдвое и рассчитывалось новое $vrf1$ до тех пор, пока не удовлетворялось неравенство

$$vrf1 < vrf0 \quad (24)$$

Значения параметров \vec{x} , при которых выполняется неравенство (24), выбирались как новые значения параметров. Контролировалось также число уменьшений шага. Если число уменьшений шага, при котором невязка не уменьшается, превышает это число, то итерационный процесс прекращается.

Таким образом, мы всегда "двигались" в "сторону" лучшего удовлетворения уравнений связи. На каждой итерации рассчитывалось так же, как и в FUMILI^[4], максимальное из отношений приращений параметра к его ошибке - ак. Итерационный процесс считался сходящимся, если $\text{ак} \rightarrow 0$. Наблюдалась чёткая корреляция между ак и невязкой (23): при $\text{ак} \rightarrow 0$ всегда $\text{vrfo} \rightarrow 0$. Итерационный процесс прекращался, когда $\text{ак} < \text{eps}$. В нашем случае $\text{eps} = 0,04$.

Литература

1. J.Peter Berge, Frank T.Solmitz and Horace D.Taft, Kinematical Analysis of Interaction Vertices from Bubble Chamber Data, The Review of Scientific Instruments, 32, No5, 1961.
2. R.Bock, Application of a Generalized Method of Least Squares for Kinematic Analysis of Tracks in Bubble Chambers, CERN 60-30(1960).
3. В.И.Мороз, Сведение минимизации квадратичного функционала с условиями связи к минимизации функционала с дополнительным квадратичным членом, ОИЯИ,Р-1958,Дубна,1965.
4. Статистические методы в экспериментальной физике, Москва, Атомиздат, 1976.
5. В.Н.Болотов и др., ЯФ 45 (1987), 1652.

Рукопись поступила в издательский отдел
6 ноября 1991 года.