

МАТЕМАТИЧЕСКИЕ ПРОБЛЕМЫ КАЛИБРОВКИ АВТОМАТИЗИРОВАННЫХ ИЗМЕРИТЕЛЬНЫХ СИСТЕМ ДЛЯ ОПТИЧЕСКИХ ТРЕКОВЫХ ДЕТЕКТОРОВ В ФИЗИКЕ ВЫСОКИХ ЭНЕРГИЙ

Н. Б. Богданова

Объединенный институт ядерных исследований, Дубна

В. Гаджоков

Институт ядерных исследований и ядерной энергетики
Болгарской АН, София

Г. А. Ососков

Объединенный институт ядерных исследований, Дубна

В обзоре рассматриваются математические проблемы, возникающие при калибровке автоматизированных систем измерения данных оптических трековых детекторов в физике высоких энергий. Дана формулировка задачи калибровки и обсуждаются порождаемые ею математические модели. Описаны этапы калибровки и применение их результатов к определению точности и стабильности систем. Приведены краткие сведения о разработанных в ОИЯИ алгоритмах и программах калибровки.

A review is presented of the mathematical problems arising when automated readout systems are being calibrated for data from optical track detectors in high-energy Physics. The problem of calibration is formulated and generated thereby mathematical models are discussed. Both the stages of calibration and the application of their results to the determination of system accuracy and stability are described. Brief information is given about calibration algorithms and computer programs developed in the JINR.

ВВЕДЕНИЕ

Одним из актуальных направлений автоматизации научных исследований является применение автоматизированных измерительных устройств для ввода в ЭВМ больших потоков экспериментальной информации.

Основной источник экспериментальных данных в физике высоких энергий — различные детекторы элементарных частиц, среди которых широко распространены оптические трековые детекторы, такие, как пузырьковые, искровые и стримерные камеры. Благодаря визуализации событий, происходящих в этих камерах, стереопроекции событий можно зафиксировать на фото пленке, как это делается

в наиболее распространенной фильмовой методике, либо занести непосредственно в память ЭВМ после оцифровки и кодирования изображения при бесфильмовом телевизионном съеме.

Выступая как долговременный носитель информации, фотопленка дает наиболее полную картину события, позволяющую очень точно реконструировать его топологию и определять кинематические параметры по кривизне в магнитном поле и ионизационной плотности треков.

Массовый поток стереоснимков, достигающий в некоторых экспериментах 10^5 — 10^6 снимков в год, сделал фильмовую методику объектом интенсивного вторжения автоматизации сначала в обработку полученных данных, а затем и в сам процесс измерения фотографий. С этой целью был разработан ряд измерительных автоматов, осуществляющих под управлением мини-ЭВМ сканирование камерных снимков. В соответствии со способами сканирования, т. е. просмотра снимка световым лучом, эти устройства можно классифицировать на несколько групп.

К первой из них относятся устройства с глобальным просмотром всего поля сканирования, они основаны на принципе механической развертки и отличаются только ее траекторией и видом светового пятна. Устройство Хафа и Пауэлла — HPD [1] или аналогичные отечественные приборы [2] осуществляют параллельную развертку пятном круглой формы. Спиральный измеритель (СИ) [3, 4] сканирует изображение по спирали радиально ориентированной узкой щелью от вершины события, измеряемой оператором, для чего в конструкцию СИ, помимо основной полярной, введена вторая, прямоугольная система координат для ручных измерений.

Другая группа устройств выполняет электронно-лучевое сканирование, гибкость которого позволяет совмещать сканирование с поиском вершин и треков, осуществляя растровую развертку отдельных участков снимка. Сюда относятся установки PEPR, POLLY, ERASME, MEJAC, АЭЛТ-2 и др., описания которых можно найти в обзорах [1, 5] и трудах конференций [6—8]. Отметим еще устройства типа SWEEPNIK [6], осуществляющие аппаратное слежение за треком с помощью вращающейся щели, наводимой на трек оператором. Аналогично работает система HEVAS (см. [8, с. 326]).

Для сокращения потока информации, поступающей с автоматических сканаторов (особенно с тех, что относятся к первой группе), обычно осуществляется селекция данных в процессе сканирования с помощью целеуказаний, полученных заранее, при предварительном просмотре снимков, и измерения некоторых точек на них. Это делают на специальных просмотрно-измерительных столах, работающих в полуавтоматическом режиме (ручные измерения с автоматической оцифровкой, проверкой и сбором данных на магнитной ленте). Еще острее проблема быстрого сокращения потока информации встает в системах бесфильмового телевизионного съема данных с искровых и стримерных камер [9, 10], где этот поток благодаря высокой частоте

те срабатывания камер достигает 10^6 — 10^7 бит/с. Бесфильмовый съем позволяет повысить скорость регистрации событий, исключить трудоемкий и дорогостоящий этап измерения фотографий, осуществить контроль за ходом эксперимента и качеством регистрируемых событий. Проблема сжатия данных при бесфильмовом их съеме ставится совсем иначе, чем в фильмовой методике. Сжатие в телевизионных системах должно принципиально выполняться в реальном времени, т. е. в темпе телевизионной развертки, и при этом из-за невозможности перемеров не должно допускать информационных потерь. Для преодоления противоречивости этих требований в [11, 12] были предложены быстрые алгоритмы сжатия, ориентированные на реализацию в виде специальной аппаратуры сжатия данных (АСД). Оцифрованное изображение одной из проекций события, состоящее из 1—2 тыс. точек АСД, сжимает до 1—2 сотен линейных трек-элементов (ЛТЭ).

Как сканирующие автоматы и системы бесфильмового съема, так и просмотрово-измерительные столы выдают данные измерений в виде упорядоченных пар чисел, представляющих измеряемые точки снимка в системе координат, задаваемой отсчетными устройствами прибора.

Поскольку для входа в программы обработки результаты измерений должны быть представлены в некоторой унифицированной системе декартовых координат (в качестве которой обычно принимается прямоугольная система координат, связанная с изображением), необходимо преобразовывать данные оцифровки, выдаваемые прибором при измерении, к этой унифицированной системе координат. Системы координат, в которых работают отсчетные устройства прибора, зависят от его конструкции и могут быть прямоугольными (таких большинство, полярными (как у спирального измерителя) и даже биполярными (как, например, у черновских столов системы MYLADY). Математическим аспектам проблемы установления прямой и обратной зависимости между системами координат самого прибора и унифицированных координат, связанных с изображением, и посвящена настоящая работа.

После постановки задачи и введения необходимого математического аппарата дальнейшее изложение будет вестись в соответствии с основными этапами обработки калибровочных измерений.

В разд. 3 излагаются методы определения координат центров крестов по данным сканирования калибровочной решетки. С этой целью вначале каждый крест должен быть распознан среди данных сканирования либо как целая фигура, либо (что чаще всего) как два отрезка прямых, его составляющих. Поскольку заранее известны приблизительное расположение каждого креста и его форма (т. е. угол раствора плеч, обычно равный 90°), то задача распознавания упрощается. Однако в более общей постановке, учитывающей возможные значительные искажения оптической и электронной систем сканатора и неортогональность сканирования, приходится приме-

нять более сложные методы, которые будут подробно изложены в третьем разделе работы.

После нахождения массивов точек, составляющих плечи креста, их аппроксимируют прямыми или при необходимости полиномами подходящей степени с помощью подгонки по методу наименьших квадратов (МНК) с выбросом далеко отстоящих точек или более надежными, так называемыми робастными методами подгонки.

Решая совместно приближенные уравнения плеч, определяют координаты центров крестов калибровочной решетки и среднеквадратичные погрешности этих координат. На основе этих данных на следующем этапе обработки осуществляется построение прямых и обратных калибровочных преобразований, а также определяются точностные характеристики прибора. Основным математическим инструментом этого этапа являются системы двумерных ортогональных полиномов.

Проблема построения и ортогонализации системы двумерных полиномов, удобных с вычислительной точки зрения, прошла свою эволюцию в их приложениях к прямым и обратным калибровочным преобразованиям.

В одной из ранних работ по калибровке PEPR [13] была использована идея построения системы двумерных полиномов перемножением одномерных многочленов Чебышева.

Более развернутое исследование систем двумерных полиномов специально для целей построения прямого и обратного калибровочного преобразования, а также выбора степени таких полиномов, оптимальной с точки зрения достигаемой точности аппроксимации, было дано в [14, 15]. Более эффективные по скорости вычислений системы полиномов были предложены в недавних работах, посвященных двумерным [16] и трехмерным [17] полиномам. Сводка этих результатов будет дана в разд. 4. Методы повышения точности калибровочных преобразований в областях между узлами калибровочной решетки описаны в разд. 5, где также даны методы исследования стабильности работы измерительных приборов. Описан способ использования приборов, имеющих дрейф калибровочных параметров в случаях, когда этот дрейф носит регулярный характер, как, например, вследствие прогрева прибора в течение нескольких часов после его включения.

Раздел 6 настоящей работы посвящен проблемам программной реализации алгоритмов калибровки и обзору имеющихся стандартных калибровочных программ.

1. ПОСТАНОВКА ЗАДАЧИ

Представим изображение в виде двумерного набора точек, подлежащих измерению. Тогда с каждой точкой C в плоскости изображения можно связать две пары чисел: ее декартовы координаты (x, y) в этой плоскости и пару целых чисел (u, v) , получаемых с отсчетных устройств прибора в результате акта измерения точки C .

Возникает задача калибровки: установить вид калибровочного преобразования $(u, v) \xrightarrow{F} (x, y)$, позволяющего правильно интерпретировать данные измерений, а также определить, насколько можно им доверять, т. е. найти точностные характеристики прибора во всех точках поля измерения и убедиться в их стабильности во времени.

Виды калибровочных эталонов. Основой методики калибровки является измерение специального эталонного изображения, представляющего набор объектов, пригодных для автоматических измерений, размещенных равномерно по всему полю сканирования с известными с высокой степенью точности (~ 1 мкм) значениями координат, что позволяет отождествить их с «идеальными» координатами (x, y) .

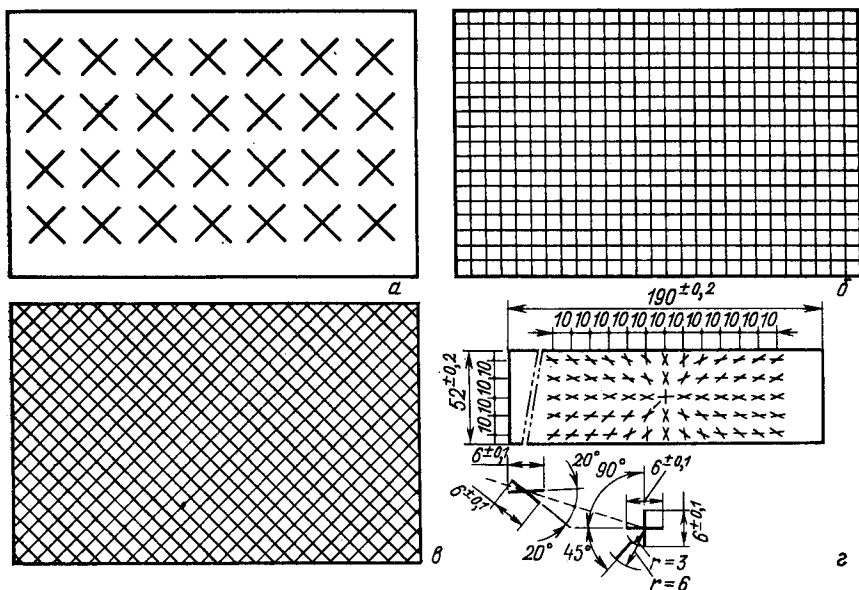


Рис. 1. Калибровочные решетки разных типов

Таким объектом может служить фигура в форме креста с известными координатами центра, т. е. точки пересечения плеч этого креста. При правильном соотношении длины плеч и шага сканирования с каждого плеча будет получено 15—20 отсчетов, что достаточно для оценки координат центра креста. Как показано в обзоре [5], для обеспечения одинакового разброса отсчетов от каждого плеча крест должен быть сориентирован так, чтобы линия сканирования пересекала оба плеча под одинаковыми углами. Оптимальность угла раствора креста с точки зрения получения равных ошибок центра по осям достигается при 90° . Размещение центров крестов с равным шагом по всему полю сканирования дает калибровочную решетку, кресты которой могут быть либо отделены один от другого (рис. 1, а), либо

образуют сетку из прямых линий (рис. 1, б). Для ручных измерений, а также для электронно-лучевого сканирования на автоматах типа PEPR с микрорастром произвольной ориентации и световым пятном в форме штриха более удобной является сетка из горизонтальных и вертикальных линий — так называемая решетка Готье (рис. 1, в). В случае приборов типа спирального измерителя (СИ), совмещающих полярную и декартову системы координат, конструкция калибровочной решетки должна позволять вести на ней измерения в обеих системах координат. Чтобы измерительная щель полярной системы координат СИ давала одинаковые сигналы от обоих плеч каждого креста, все кресты, кроме центрального, должны располагаться симметрично относительно полярных радиусов, проведенных в их центры, причем для регистрируемости этих сигналов угол между каждым из плеч и щелью не должен превышать 20° . Вид соответствующей калибровочной пластины, используемой в настоящее время для калибровки СИ ОИЯИ, приведен на рис. 1, г. Заметим, что раствор в 40° дает невыгодные условия для измерений, но он не может быть уменьшен, так как при этом резко падает радиальная точность определения центра креста. От этого недостатка свободна, например, пластина станфордского СИ [18], однако там для определения центров крестов требуется высокоточная оцифровка амплитуд сигналов, поступающих с измерительной щели (заметим, что, например, в конструкции СИ ОИЯИ это не предусмотрено).

Этапы обработки. После проведения калибровочных измерений проводится их математическая обработка для решения задачи калибровки. Академик А. Н. Тихонов выделяет следующие основные этапы математической обработки результатов измерений на ЭВМ [19].

1. Первичная обработка, включающая нормировку данных, их фильтрацию для получения выходных результатов эксперимента.

2. Анализ установок, т. е. построение оператора F , отвечающего модели установки. Анализ проводится: а) постановкой специальных юстировочных экспериментов; б) математическим (численным) моделированием.

3. Интерпретация результатов.

В общей цепи обработки данных автоматических измерений калибровка олицетворяет второй этап. Определение калибровочного преобразования на основании юстировочных измерений и означает построение оператора, характеризующего влияние измерительной установки в математических терминах.

Можно, однако, само калибровочное измерение рассматривать как некоторый эксперимент. При таком подходе схема А. Н. Тихонова определяет этапы обработки калибровочного эксперимента.

Первый этап. Распознавание и фильтрация крестов и определение их центров.

Второй этап. Построение калибровочного преобразования. Отметим здесь пользу имитационного моделирования для анализа прибора и проверки алгоритмов обработки калибровочных данных. О мето-

дах такого моделирования будет более подробно говориться ниже в разд. 2.

Третий этап. Интерпретация результатов, т. е. определение обоснованности модели, послужившей для вывода калибровочного преобразования, и точностных характеристик прибора.

Как показала практика [20, 21], для проверки качества калибровочных коэффициентов полезно провести дополнительное измерение простого, но очень точно выполненного эталонного изображения (обычно эта прямая линия, проведенная с гарантированным отсутствием искривлений и отклонений, больших ± 1 мкм).

Простейшие модели. Чтобы более наглядно представить постановку задачи об определении калибровочного преобразования, начнем с самого простого случая. Рассмотрим идеальный прибор без дисторсий с прямоугольной системой координат (u, v) , ортогональность осей которых гарантирована. Даже в таком простом случае преобразование $(u, v) \xrightarrow{F} (x, y)$ требует задания пяти параметров: двух значений масштабов по осям M_x и M_y , двух чисел x_0, y_0 , задающих начальные положения отсчетных систем по осям, и угла φ между соответствующими осями в обеих системах координат. Для определения этих параметров недостаточно просто измерить пять точек с известными идеальными координатами, так как всякое реальное измерение производится с неизвестной случайной погрешностью, проекции которой на оси обозначим Δ_x, Δ_y . Получаем самую простую модель для описания соотношения $(u, v) \xrightarrow{F} (x, y)$:

$$\left. \begin{aligned} x &= x_0 + uM_x \cos \varphi + vM_y \sin \varphi + \Delta_x; \\ y &= y_0 - uM_x \sin \varphi + vM_y \cos \varphi + \Delta_y. \end{aligned} \right\} \quad (1)$$

Следует сразу же отметить два обстоятельства, осложняющие использование соотношения (1): нелинейность модели по параметрам и наличие небольшой неортогональности осей практически во всех реальных измерительных системах.

Учет малой косоугольности осей обеспечивает более общая, чем (1), аффинная зависимость, являющаяся в то же время и более удобной в вычислительном отношении, благодаря линейности по параметрам

$$\left. \begin{aligned} x &= x_0 + au + bv + \Delta_x; \\ y &= y_0 + cu + dv + \Delta_y. \end{aligned} \right\} \quad (2)$$

Поэтому соотношение (2) остается предпочтительным, даже если после оценки шести параметров окажется, что условие ортогональности осей

$$ac + bd = 0 \quad (3)$$

выполняется с высокой степенью точности. Более наглядные по смыслу параметры из соотношения (1) могут быть легко вычислены по a , b , c , d :

$$M_x = \sqrt{a^2 + c^2}; \quad M_y = \sqrt{b^2 + d^2}; \quad (4)$$

$$\varphi = \arctg(ab - cd)/2ad \quad (5)$$

(см. также [22]). Как известно [23], для исключения случайных погрешностей типа Δx , Δy число уравнений должно значительно превышать число неизвестных, что дает возможность применения статистических методов типа МНК. Коэффициент косоугольности в явном виде, как разность прямого угла и угла между осями прибора, также находится из параметров (2):

$$\kappa = \arcsin(ac + bd) / \sqrt{(a^2 + b^2)(c^2 + d^2)}. \quad (6)$$

Преобразование (2) остается удобным и в самом простом случае равных масштабов по осям $M_x = M_y$, когда соотношения (1) и (2) зависят фактически всего от четырех параметров. Проверка близости значений масштабов, вычисляемых по формулам (4), и выполнения условия (3) в этом случае может служить хорошим тестом для контроля проводимых вычислений.

Преобразование (2) линейно не только по параметрам, но и в отношении входящих в него переменных u , v . Последнее обстоятельство явилось следствием сделанного в самом начале предположения об отсутствии дисторсий и нуждается в проверке даже для таких простых не автоматических приборов для ручных измерений, как микроскоп УИМ или просмотрово-измерительный стол БПС-2 [22]. Реальные измерительные автоматы являются сложными системами из оптических, осветительных узлов, механических или электронных устройств, выполняющих развертку сканирующего пятна. Каждый из этих узлов обладает определенными дисторсиями и даже при самой тщательной юстировке может внести заметные нелинейные искажения в снимаемые координаты. Эти нелинейности могут быть по-разному учтены в калибровочном преобразовании. Один из подходов был уже намечен в преобразовании (1), коэффициенты которого оказались параметрами отсчетных систем прибора. Такой путь введения в нелинейное калибровочное преобразование коэффициентов, явно учитывающих дисторсии, дает прекрасную возможность быстрой диагностики источников этих дисторсий по факту превышения каким-то из коэффициентов заданного для него предела. Реализация подобного подхода в калибровке спирального измерителя будет изложена ниже. Однако, несмотря на привлекательность этого пути для инженеров, разрабатывающих и эксплуатирующих устройство, он не может рассматриваться как общий, так как требует для каждого измерительного прибора строго индивидуального подхода, основанного на детальном анализе оптических, кинематических и электронных схем, и вдобавок ведет к сложным с вычислительной точки зрения задачам

минимизации нелинейных функционалов для определения искомых параметров.

Линейные аппроксимационные модели. Более универсальным является подход с использованием линейных по параметрам аппроксимирующих функций вида

$$F(u, v, A) = \sum_{j=1}^M a_j \varphi_j(u, v), \quad (7)$$

где $\{\varphi_j(u, v)\}$ — некоторая система линейно-независимых непрерывных функций двух переменных; a_j — компоненты неизвестного вектора параметров A , связь которых с дисторсиями конкретных узлов прибора не обязательна (хотя при необходимости, как правило, такую связь с некоторыми комбинациями этих параметров удастся установить). Такой подход позволяет применить всю мощь хорошо разработанных линейных регрессионных методов математической статистики для оптимальной оценки векторов A и B , определяющих преобразования

$$\hat{x} = F(u, v, A), \quad \hat{y} = F(u, v, B) \quad (8)$$

с F вида (7).

Задача получения калибровочного преобразования, наилучшим образом соответствующего проведенным калибровочным измерениям, может быть поставлена следующим образом.

Пусть калибровочная решетка состоит из N крестов, идеальные координаты центров которых $\{x_k, y_k\}_{k=1}^N$. Обработка данных автоматического сканирования этой решетки дает измеренные величины этих координат $\{u_k, v_k\}_{k=1}^N$ с матрицей ошибок S_{2N} , которая в общем случае может учитывать корреляции между измерениями обеих координат в разных точках, т. е. $\text{cov}(u_k, v_j) \neq 0$; $j, k = \overline{1, N}$. Требуется по этим данным калибровочного измерения найти наилучшие оценки A_N и B_N параметров, определяющих преобразования (8).

О качестве преобразования судят по близости точек (x, y) и (\hat{x}, \hat{y}) , причем эта близость должна быть одинаково справедлива во всех точках поля измерений. Имея N измеренных точек, равномерно расположенных на поле, мы можем оценить меру близости по малости абсолютных величин остатков (невязок) преобразований (8):

$$dx_k = x_k - F(u_k, v_k, A_N); \quad dy_k = y_k - F(u_k, v_k, B_N), \quad (9)$$

образующих вектор $D_{2N} = \text{col}(dx_1, \dots, dx_N, dy_1, \dots, dy_N)$.

При достаточно плотном расположении крестов (т. е. достаточно большом N) выбор размерности преобразования M и таких параметров A_N, B_N , которые обеспечивают малость абсолютных величин всех остатков, фактически означает, что в преобразованиях (8) удалось учесть все систематические отклонения, порождаемые дисторсиями

прибора. Поэтому так называемая карта остатков (рис. 2), образованная увеличенными в $10^3 - 10^4$ раз двумерными векторами $\{dx_k, dy_k\}$, выходящими из точек $\{x_k, y_k\}$, является весьма полезным и наглядным средством для оценки правильности выбора параметров.

В качестве меры близости точек (x, y) и (\hat{x}, \hat{y}) естественно взять среднеквадратичную величину остатков, усредненную с учетом их погрешностей, т. е.

$$L_2^2 = \sqrt{D_{2N}^T W_{2N} D_{2N}},$$

где T означает транспонирование, $W_{2N} = S_{2N}^{-1}$.

Как известно [23], в предположении нормальности распределения вектора остатков D_{2N} оптимальные оценки параметров A_N, B_N могут быть получены минимизацией функционала L_2 . Если справедливо дополнительное предположение о независимости остатков, то S_{2N} становится диагональной и задача фактически распадается на две независимых задачи вычисления оценок A_N и B_N минимизацией функционалов, имеющих с учетом (7) — (9) вид:

$$\left. \begin{aligned} \Phi_x^2 &= \sum_{k=1}^N w_{xk} \left[x_k - \sum_{j=1}^M a_j \varphi_j(u_k, v_k) \right]^2; \\ \Phi_y^2 &= \sum_{k=1}^N w_{yk} \left[y_k - \sum_{j=1}^M b_j \varphi_j(u_k, v_k) \right]^2 \end{aligned} \right\} \quad (10)$$

с весами

$$w_{xk} = \sigma_{u_k}^{-2}, \quad w_{yk} = \sigma_{v_k}^{-2}; \quad k = \overline{1, N}.$$

Очевидна важность правильной параметризации задачи калибровки, т. е. удачного выбора системы двумерных функций $\{\varphi_j\}$, которые, с одной стороны, должны достаточно хорошо описывать встречающиеся на практике типы дисторсий, а с другой — обеспечивать вычислительные удобства при решении системы нормальных уравнений (быстрая вычислимость на ЭВМ, недопущение ситуаций плохой обусловленности и т. д.).

Есть еще один аспект проблемы параметризации. Преобразования (8) дают нам прямой переход $(u, v) \xrightarrow{F} (x, y)$ от измеренных координат к идеальным. Однако в практике измерений возникает необходимость в обратном преобразовании $(x, y) \xrightarrow{F^{-1}} (u, v)$, обеспечивающем, в частности, привязку некоторых заранее измеренных вручную точек к данным автоматических измерений, что необходимо для организации быстрой селекции данных в процессе сканирования.

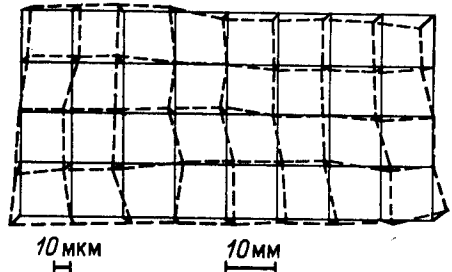


Рис. 2. Пример карты остатков

К сожалению, нелинейность системы $\{\varphi_j(u, v)\}_{j=1}^M$ препятствует непосредственному обращению (8), поэтому для получения функции обратного преобразования $F^{-1}(x, y, A')$ ее также целесообразно построить в виде линейной комбинации двумерных функций, образующих систему $\{\psi_j(x, y)\}_{j=1}^M$:

$$F^{-1}(x, y, A') = \sum_{j=1}^M a_j' \psi_j(x, y). \quad (11)$$

Коэффициенты (11) также можно определить по МНК, используя данные измерений $\{u_h, v_h\}$ и идеальные координаты решетки $\{x_h, y_h\}_{h=1}^N$, но в отличие от прямого преобразования минимизируемый функционал в случае обратного преобразования строится на фиксированном наборе точек $\{x_h, y_h\}_{h=1}^N$, что значительно упрощает вычислительную сторону проблемы.

Переходя к проблеме конструирования систем функций $\{\varphi_j\}$ и $\{\psi_j\}$, укажем на один из возможных путей такого конструирования, приведенный ниже в разделе о калибровке для прибора с полярной системой координат. Функции $\{\varphi_j\}$ там не являются ортогональными, поскольку строятся линеаризацией по параметрам соотношений, учитывающих аппаратные особенности и дисторсии сканатора. Подобный опыт хотя и носит частный характер, но достаточно поучителен, так как показывает, как могут быть установлены такие соотношения функций $\{\varphi_j\}$, которые позволяют осуществлять диагностику прибора.

Однако чаще всего в качестве $\{\varphi_j\}$ и $\{\psi_j\}$ используются системы двумерных ортогональных полиномов. Они особенно удобны при калибровке приборов наиболее общепотребительного типа с прямоугольной системой координат, так как упрощают вычисления и позволяют хорошо учесть самые разнообразные дисторсии, в том числе значительные подушкообразные или бочкообразные искажения, присущие обычно приборам с электронно-лучевым сканированием, а также бесфильмовым системам. Более подробное описание различных систем полиномов дано в следующем разделе в числе других математических сведений, необходимых для реализации этапов обработки.

2. МАТЕМАТИЧЕСКИЙ АППАРАТ

Имитационное моделирование. Один из наиболее плодотворных путей проверки правильности как самой математической модели явления, зависящего от случайных параметров, так и нашей способности разумно интерпретировать результаты этого явления — это имитационное моделирование. Метод Монте-Карло позволяет заложить в модель поток данных, поступающих с автомата при сканировании калибровочной решетки, случайные погрешности, появляющиеся в процессе сканирования, и систематические отклонения вследствие

учтенных дисторсий прибора. Статистическая обработка такой модели с помощью системы программ, реализующей некоторые принципы и алгоритмы, позволяет: а) проверить правильность алгоритмов и их способность к обнаружению систематических погрешностей; б) определить точность вычисляемых параметров; в) установить характеристики скорости и надежности и провести верификацию всей системы программ. Для обеспечения высокого правдоподобия статистических выводов моделирование и последующая обработка должны быть повторены большое число ($10^3 - 10^4$) раз.

Рассмотрим в качестве примера модель симметричной калибровочной решетки, состоящей из N отдельно расположенных крестов с взаимно перпендикулярными плечами и идеальными координатами центров $\{x_k, y_k\}_{k=1}^N$ (см. рис. 1, а).

За начало идеальной системы координат удобно принять центр симметрии решетки, и поскольку начало координат отсчетных систем прибора должно находиться за пределами поля сканирования, то оценками коэффициентов переноса \hat{x}_0, \hat{y}_0 между этими системами могут служить средние арифметические координат всех точек, полученных при сканировании решетки. Конструкция прибора, а также данные предыдущих или контрольных измерений позволяют сделать начальные оценки масштабов по осям: M_{x0}, M_{y0} .

Этих априорных данных достаточно для приблизительного определения границ квадратной области, содержащей k -й крест. Ее центр имеет координаты

$$U_{k0} = x_k M_{x0} + \hat{x}_0, V_{k0} = y_k M_{y0} + \hat{y}_0; k = \overline{1, N}, \quad (12)$$

а сторона равна $L/\sqrt{2}$, где L — длина каждого из плеч креста.

Это определяет обычную процедуру первого этапа обработки: массив данных сканирования всей калибровочной решетки разбивается на N подмассивов, образованных точками, попавшими в области, относящимися к каждому кресту, которые и обрабатываются по отдельности. В соответствии с такой схемой обработки можно осуществить следующую простую имитацию процесса сканирования калибровочной решетки.

Перенесем область, связанную с k -м крестом, в начало координат. Уравнения плеч креста запишется в виде

$$\left. \begin{aligned} v &= (\lambda + \Delta_k) u + \Delta_{1k}, \\ v &= -(\lambda - \Delta_k) u + \Delta_{2k} \end{aligned} \right\} \quad (13)$$

при ограничении $|u| \leq L/\sqrt{1 + \lambda^2}$ (взаимная перпендикулярность плеч означает, что $\lambda = 1$). Здесь Δ_{1k} и Δ_{2k} — отрезки, отсекаемые плечами на оси v , а Δ_k — возможное смещение направления плеч. Появление этих трех параметров вызвано неточностью априорных

знаний и возможными дисторсиями прибора, которые предполагаются малыми, чтобы обосновать сохранение прямолинейности плеч креста. Задаваясь шагом сканирования h , дискретизируем ось u : $u_i = ih$; $i = 0, \pm 1, \dots, \pm n$; $n = [L/\sqrt{1 + \lambda^2}]$ и из (13) для каждого u_i получаем два значения v , соответствующие двум плечам креста.

Для имитации погрешностей отсчетных систем к каждой из координат следует добавить нормально распределенную величину с нулевыми средними и стандартными отклонениями σ_u, σ_v . Чтобы учесть возможность сбоев отсчетных систем, каждая из точек при записи в массив может быть пропущена (не записана) с малой вероятностью $P_{\text{проп}} \sim 1/20$. Кроме этого, к массиву данных сканирования креста добавляются шумовые отсчеты, для чего при каждом u_i с вероятностью $P_{\text{шум}}$ помимо двух значений v , полученных из уравнений (13), разыгрывается еще одно значение v , равномерно распределенное в отрезке $(-L\lambda/\sqrt{1 + \lambda^2}, +L\lambda/\sqrt{1 + \lambda^2})$.

Если требуется промоделировать какую-то конкретную картину дисторсий (например, подушкообразных), то исходя из этой картины составляется карта дисторсий, т. е. смещений по осям $\{du_k, dv_k\}$ (в отсчетных единицах) во всех N узлах решетки, а также зависящая от них таблица Δ_k — отклонений в ориентировке крестов. Величины Δ_{1k} и Δ_{2k} для уравнений (13) вычисляются по очевидным формулам: $\Delta_{1k} = dv_k + (\lambda + \Delta_k) du_k$, $\Delta_{2k} = dv_k - (\lambda - \Delta_k) du_k$.

Теперь если потребуется модель полного массива данных, полученных при сканировании всей калибровочной решетки, то следует добавить к координатам всех точек каждого из подмассивов, относящихся к отдельному кресту, приблизительные координаты центров этих крестов из (12), слить все эти подмассивы в один и расположить их при каждом u в порядке возрастания v .

При необходимости аналогичным образом может быть смоделировано и множество связанных отсчетов, возникающих при появлении на калибровочной решетке царапин, пылинок и т. д.

Гораздо более сложной получается модель при спиральном сканировании, так как там пересечения прямых, соответствующих плечам крестов и спирали, приводят к трансцендентным уравнениям. Кроме того, много сложнее разыгрываются координаты шумовых точек, которые должны быть разбросаны равномерно по длине дуги спирали. Реализация этой модели, описанная в [24, 25], позволила провести отладку калибровочной программы еще на стадии разработки СИ, предложить и отладить новые алгоритмы измерения реперных крестов и подтвердить правильность (верифицировать) программы.

2. Робастные оценки в регрессионных задачах. Рассмотрим регрессионную зависимость $y = \sum_j x_j b_j + d$, где x_j — факторы, определяемые местом, где производится измерение, y — отклик (измеряемая величина), d — погрешность измерения, имеющая нулевое среднее и дисперсию σ^2 , b_j — неизвестные регрессионные коэффициенты,

$j = \overline{1, m}$. Измерения осуществляются независимо в n точках, что дает

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Тогда справедливо $Y = XB + D$, где $B = \text{row}(b_1, \dots, b_m)^T$; $D = \text{row}(d_1, \dots, d_m)^T$ — неизвестные векторы (T означает транспонирование). Случайный вектор D имеет скалярную ковариационную матрицу $\text{cov } D$.

Известно [23], что когда величина d распределена по нормальному закону, оптимальной оценкой вектора параметров является оценка по методу наименьших квадратов (МНК):

$$\Phi(b_1, \dots, b_m) = \sum_i w_i (y_i - \sum_j x_{ij} b_j)^2 \rightarrow \min. \quad (14)$$

МНК — оценка B линейно выражается через Y :

$$\hat{B} = (X^T W X)^{-1} X^T W Y. \quad (15)$$

Здесь W — матрица весов, которая при независимых наблюдениях имеет диагональный вид, $w_{ii} = \sigma_i^{-2} = w_i$ (σ_i^2 — дисперсия i -го наблюдения).

Для наглядности распишем более подробно матричную запись (15) для случая $m = 2$ (прямая линия: $y = ax + b$), который будет неоднократно использоваться в дальнейшем.

Для определения двух параметров a и b минимизируем функционал

$$\Phi(a, b) = \sum_{i=1}^n w_i (y_i - ax_i - b)^2 = \sum_{i=1}^n w_i d_i^2. \quad (16)$$

С этой целью приравниваем к нулю производные $\Phi(a, b)$ и получаем систему нормальных уравнений, которая обычно записывается с использованием так называемых гауссовых скобок $[wx^p y^q] = \sum_{i=1}^n w_i x_i^p y_i^q$:

$$\left. \begin{aligned} a [wx^2] + b [w x] &= [wxy]; \\ a [wx] + b [w] &= [wy]. \end{aligned} \right\} \quad (17)$$

Детерминант системы (17) $\det = [wx^2] [w] - [wx]^2$; оценки параметров (т. е. компоненты вектора \hat{B} в двумерном случае):

$$\left. \begin{aligned} \hat{a} &= [wxy] [w] - [wy] [wx] / \det; \\ \hat{b} &= [wx^2] [wy] - [wxy] [wx] / \det \end{aligned} \right\} \quad (18)$$

являются случайными величинами со средними a и b и матрицей ковариации

$$\text{cov } \hat{B} = \sigma^2 (X^T W X)^{-1} = \sigma^2 \frac{1}{\det} \begin{pmatrix} [w] & -[wx] \\ -[wx] & [wx^2] \end{pmatrix},$$

так что

$$\begin{aligned} \sigma_a &= \sigma \sqrt{\frac{[w]}{\det}}, \quad \sigma_b = \sigma \sqrt{\frac{[wx^2]}{\det}}, \\ \rho_{\hat{a}, \hat{b}} &= -\frac{[wx]}{\sqrt{[w][wx^2]}}. \end{aligned} \quad (19)$$

Оценка σ может быть получена как

$$\hat{\sigma} = \sqrt{\frac{\Phi_{\min}(a, b)}{n-2}} = \sqrt{\frac{\sum_i w_i \hat{d}_i^2}{n-2}},$$

где остаток (невязка) в i -й точке обозначен $\hat{d}_i = y_i - \hat{y}_i = y_i - \hat{a}x_i - \hat{b}$.

Для задач выделения образов по данным автоматического сканирования характерно наличие отсчетов, порожденных шумовыми образованиями или сбоями прибора. Эти засоряющие отсчеты придают данным, по которым ведется оценка неизвестных параметров, характер неоднородной совокупности и затрудняет применение описанных выше регрессионных методов статистики, опирающихся на нормальность распределения всех данных. Поэтому для устранения этого засорения используются различные способы так называемого выброса. Почти все они сводятся к МНК — оценке регрессионных параметров по всем данным с проверкой критерия $\hat{\sigma} < \sigma_{\text{lim}}$ (σ_{lim} — задаваемая константа, обычно $\sigma_{\text{lim}} \sim 10 \div 20$ мкм). При нарушении критерия производится выброс, т. е. исключение тех точек, для которых $|\hat{d}_i| > 3\hat{\sigma}$, или в более надежном варианте [26] — исключение $\max_i |d_i|$. После этого оценки \hat{a} , \hat{b} , $\hat{\sigma}$ находятся снова по урезанной совокупности. Критерием прекращения выброса служит малость $\hat{\sigma}$.

Громоздкость и недостаточная надежность подобных процедур привели к все более широкому использованию так называемых робастных (устойчивых) способов оценки параметров неоднородных совокупностей.

Суть их заключается в замене квадратичного функционала (14), (16) функционалом с более общей функцией вклада $\psi(t)$:

$$\Phi_{\psi}(b_1, \dots, b_m) = \sum_i \psi\left(\frac{\hat{d}_i}{s}\right), \quad (20)$$

где s — оценка параметра масштаба, равная σ , если σ известна.

Выбор функции вклада определяется степенью засорения и удобством вычислений. Он направлен на снижение влияния посторонних, засоряющих точек, т. е. на уменьшение вклада в функционал тех точек, отклонения которых $|\hat{d}_i|$ велики.

Подробные обзоры робастных методов можно найти в [27, 28], более краткие сведения с примерами приложений к обработке трековых данных приведены в сообщении [29].

Здесь мы ограничимся упоминанием одних типов функций вклада — выпуклой, предложенной Хьюбером [28]

$$\psi_1(t) = \begin{cases} t^2/2, & |t| < c; \\ c|t| - t^2/2, & |t| \geq c, \end{cases}$$

и невыпуклой Тьюки [30]

$$\psi_2(t) = \begin{cases} \frac{c^2}{6} + \frac{(t^2 - c^2)^3}{6c^4}, & |t| < c; \\ c^2/6, & |t| \geq c. \end{cases} \quad (21)$$

Метод Хьюбера корректен (функционал Φ_{ψ_1} имеет единственный минимум) и оптимален в минимаксном смысле [27]. Однако он дает удовлетворительные оценки лишь при не очень тяжелых «хвостах» плотности вероятности отклонения от среднего (коэффициент эксцесса не более 5—6). Для случаев с большей засоренностью данных лучшие результаты дают методы с невыпуклой функцией вида $\psi_2(t)$, активнее подавляющие влияние точек с большими $|\hat{d}_i|$. Платой за это является возникновение ложных локальных минимумов функционала (20).

Реализуются описанные методы в виде итерационных процедур, где на каждом шагу применяется МНК с весами $w_i = \psi'(\hat{d}_i/s)/\hat{d}_i$, вычисляемыми по оценкам параметров на предыдущих итерациях. В качестве начальных приближений параметров выбираются либо МНК — оценки с $w_i = 1$, либо априорные значения \hat{a}_0 и \hat{b}_0 , если они известны, как, например, в случае крестов калибровочной решетки. В качестве оценки параметра масштаба s используется $\hat{\sigma} = \text{med } |\hat{d}_i|/0,6745$ — нормированная медиана (средний член вариационного ряда модулей остатков). Коэффициент нормировки 0,6745 подобран так, чтобы в случае гауссова распределения $\hat{\sigma}$ была несме-

щенной оценкой. При увеличении засорения этот коэффициент следует увеличить. Константа c в формуле (21) выбирается обычно равной 5 [30]. Исследование [29] показало, что надежность робастных оценок сильно зависит от правильного выбора начальных значений параметров.

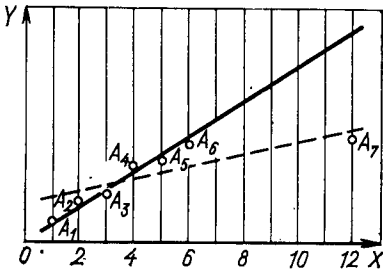


Рис. 3. Пример регрессии с резко выпадающей точкой

В некоторых вполне реальных ситуациях, когда после большого разрыва во множестве точек, по которым ведутся оценки параметров, появляется явно посторонняя точка «аутсайдер» (рис. 3), применение МНК-оценки с весами $w_i = 1$ дает значения начальных приближений \hat{a}_0 , \hat{b}_0 вблизи ложного локального минимума функционала, куда и «скатываются» оценки параметров при дальнейших итерациях (см. пунктир на рис. 3).

Как показано в [29], этой опасности удастся избежать нормировкой каждого остатка на его собственное среднеквадратичное отклонение δ_i , вычисляемое для $m = 2$ как полином второй степени от x_i с коэффициентами, найденными по данным предыдущей итерации. Фактически это означает усложнение формулы для «бвесов» Тьюки

$$w_i = [1 - (\hat{d}_i / 4\hat{\sigma}_i)^2]^2. \quad (22)$$

Системы полиномов и методы их ортогонализации. В настоящем разделе будут рассмотрены некоторые свойства линейных метрических пространств, имеющие отношение к численному построению преобразований калибровки.

Пусть Q_M — линейное действительное пространство M измерений. Пусть для любых $f, g \in Q_M$ определено скалярное произведение (f, g) , удовлетворяющее известным аксиомам [31]. Тогда Q_M является метрическим пространством, а его базис $\{b_i\}_{i=1}^M$ называется ортогональным, если выполняется условие

$$(b_i, b_j) = h_i^2 \delta_{ij} \quad (23)$$

(δ_{ij} — символ Кронекера), и ортонормированным, если все $h_i^2 \equiv 1$.

По причинам, отмеченным во введении, при аппроксимации калибровочных зависимостей нас интересуют прежде всего такие пространства Q_M , элементы которых являются полиномами от независимых переменных.

Рассмотрим полиномы от n переменных x_1, x_2, \dots, x_n (причем в физическом пространстве $n \leq 3$) в качестве элементов действительного линейного метрического функционального пространства Q_M . Начнем рассмотрение в «естественном» полиномиальном базисе, состоящем из одночленов степеней координат вида $x_1^{i_1} x_2^{i_2} \dots x_n^{i_n}$, чья

общая (суммарная) степень выражается суммой показателей $i_1 + i_2 + \dots + i_n$. Когда M — конечно, счетность элементов базиса не вызывает сомнений. При $n = 1$ элементы базиса упорядочиваются естественным образом по возрастающим степеням. Однако, если $n > 1$, существует несколько линейно-независимых одночленов одинаковой общей степени (например, при $n = 2$ имеем одночлены второй степени x_1^2 , x_1x_2 и x_2^2) и степень сама по себе оказывается недостаточной для введения порядка. Комбинаторные соображения приводят к выводу [32, 33], что существует всего

$$B_n^k = \frac{(n+k)!}{n!k!}$$

линейно-независимых полиномов от n переменных степени не выше k . В этом случае наиболее часто употребляется так называемый лексикографический порядок, который вводится следующим образом. Из двух одночленов больший номер имеет тот, у которого больше общая степень, либо при равенстве — какая-то ее часть, относящаяся к переменным с более высокими номерами.

Наличие упорядоченного базиса в Q_M позволяет рассматривать задачу о нахождении явного вида преобразования калибровки как разновидность задачи разложения заданного вектора \mathbf{F} по базису

$$\mathbf{F} = \sum_{i=1}^J a_i \mathbf{b}_i, \quad (24)$$

где $J \leq M$, $\{\mathbf{b}_i\}$ — элементы базиса в лексикографическом порядке, а $\{a_i\}$ — соответствующие (искомые) координаты вектора. Конкретные способы решения этой задачи зависят в существенной мере от выбора базиса, от определения его метрики (скалярного произведения) и от критериев оптимальности, на основе которых определяется длина J разложения (24).

Введенный естественный базис в достаточной мере элементарен и нагляден, а расчеты в нем не слишком трудоемки при использовании правила Горнера [34] или его многомерных обобщений. Однако числа обусловленности матриц*, возникающих в этом базисе, настолько высоки, что его практические применения ограничены низкими степенями независимых переменных (например, не более 5—6 при $n = 1$) даже в арифметике повышенной точности. Этого явно недостаточно для большинства задачи калибровки, что и оправдывает усилия, затрачиваемые на улучшение базиса. С теоретической точки зрения наилучшим базисом был бы тот, который обеспечивал единичную обусловленность задачи калибровки. Существование такого базиса доказано давно [35, 36], а конструктивный характер доказательства очевиден из неоднократных программных реализаций мето-

* Число или мера обусловленности матриц $\text{cond } G$ характеризует погрешность при обращении матриц. В качестве такой меры для матрицы G обычно принимают величину $\|G\| \|G\|^{-1}$ или $\lambda_{\max}/\lambda_{\min}$, где λ — характеристическое число G [31]

да [37—39]. Мы рассмотрим вкратце различные подходы к улучшению свойств базиса путем его ортогонализации и ортонормировки.

Совокупность скалярных произведений обеих сторон уравнения (24) на базисные векторы $\{b_i\}$ можно записать в матричном виде как

$$G_J \mathbf{a}_J = \mathbf{f}_J, \quad (25)$$

где $\mathbf{a}_J = \text{col}(a_1, a_2, \dots, a_J)$, $\mathbf{f}_J = \text{col}[(F, b_1), (F, b_2), \dots, (F, b_J)]$, а G_J есть матрица Грама [40] порядка J , построенная из базисных векторов. Тогда искомые коэффициенты $\{a_i\}$ выражаются в виде

$$\mathbf{a}_J = G_J^{-1} \mathbf{f}_J. \quad (26)$$

Последние два уравнения (25) и (26) делают очевидным факт непосредственного влияния выбранного базиса на обусловленность задачи калибровки.

Матричная запись (25) приводит к идее прямой ортогонализации: можно попытаться найти численно такую матрицу A , которая диагонализировала бы G_J :

$$AG_J A^{-1} = \text{diag}(g_1, g_2, \dots, g_J),$$

причем в невырожденном базисе все $g_i > 0$. Тогда

$$AG_J A^{-1} A \mathbf{a}_J = A \mathbf{f}_J$$

и компоненты вектора $A \mathbf{a}_J$ легко выразить:

$$(A \mathbf{a}_J)_i = (A \mathbf{f}_J)_i / g_i, \quad i = \overline{1, J}.$$

Способ этот трудоемок, но привлекателен обращением к богатому и хорошо известному арсеналу матричных вычислений. Он применялся с успехом для получения более точных калибровочных преобразований в [41].

Иной и более прямой путь к ортогонализации базиса до составления уравнений (25) состоит в использовании процедуры ортогонализации Грама — Шмидта [40]: если $\{b_i\}$ — исходный, а $\{e_i\}$ — ортогональный базисы, то $e_1 = b_1$ и

$$e_i = b_i - \sum_{j=1}^{i-1} \frac{(e_j, b_i)}{e_j^2} e_j, \quad i = \overline{2, J}.$$

Нетрудно убедиться, что эта процедура универсальна, т. е. она обеспечивает ортогональность набора $\{e_i\}$ не только для полиномов, но и в случае любых базисных функций $\{b_i\}$. Платой за это полезное качество является ее трудоемкость: ортогонализация по Граму — Шмидту в сочетании с нормировкой нового базиса требует расчета $J(J-1)/2$ скалярных произведений и J скалярных квадратов при размерности J , причем число необходимых скалярных произведений не зависит от размерности независимой переменной.

Третий — и на наш взгляд наиболее эффективный — способ ортогонализации полиномиального базиса основывается на использовании рекурсий типа трехчленного соотношения Форсайта [35]. Это соот-

ношение связывает значения ортогональных полиномов трех последовательных порядков от одной переменной и записывается в виде

$$P_{i+1}(x) = c_{i+1} [(x - \alpha_{i+1}) P_i(x) - \beta_i P_{i-1}(x)], \quad (27)$$

где x — произвольное, но фиксированное значение независимой переменной, а $\{\alpha_i\}$ и $\{\beta_i\}$ — рекуррентные коэффициенты, выражающиеся через скалярные произведения и не зависящие от аргумента x . Собственно соотношение Форсайта (27) касается ортогонализации базиса, но не его нормировки. Обобщения этого соотношения, сделанные в [36—39], расширили его таким образом, что:

а) была установлена связь между нормирующим коэффициентом c_i и рекуррентным коэффициентом β_i , что позволило проводить нормировку базиса параллельно с ортогонализацией;

б) соотношение было дополнено четвертым членом с тем, чтобы распространить его на все нетривиальные производные и неопределенные интегралы ортонормированных полиномов $\{P_i\}$ [38];

в) аналогичное соотношение было расписано для полиномов от нескольких переменных.

С помощью рекуррентных соотношений стало возможным численное или аналитическое построение систем ортонормированных полиномов при любом определении скалярного произведения и любом числе независимых переменных. Общая идея ортонормирующих алгоритмов этого класса состоит в следующем:

1. Предварительный расчет рекуррентных коэффициентов для всего диапазона требуемых степеней.

2. Вычисление значений полиномов базиса для произвольного фиксированного набора значений x (здесь x означает независимые переменные, необязательно в одном измерении) при помощи соответствующего рекуррентного соотношения, без того, чтобы расписать сами полиномы в коэффициентах [42].

Мы не считаем целесообразным приводить здесь явный вид полного набора расчетных формул для рекуррентных коэффициентов, которые могут быть найдены в цитированной литературе. Отметим только, что имеется взаимно однозначное соответствие между полиномами естественного и ортонормированного базисов, благодаря чему становится возможным следовать тому же лексикографическому порядку в нумерации элементов ортонормированного базиса. Нам кажется также существенным описать принцип данного подхода и аргументировать утверждение о его большей эффективности, которая к тому же является разносторонней:

рекуррентная ортонормировка требует расчета меньшего числа скалярных произведений по сравнению с методом Грама—Шмидта. Это означает, что данный подход более экономичен с точки зрения подготовительных вычислений;

рекуррентная ортонормировка требует сохранения вычисленных скалярных произведений для расчета значений полиномов, тогда как в методе Грама—Шмидта необходимо хранить значения коэффи-

циентов, чье число совпадает с числом скалярных произведений для того же метода Грама—Шмидта. Это означает, что рекуррентный подход более экономичен с точки зрения использования памяти;

при рекуррентном подходе более прост и стандартен способ расчета значений полиномов базиса, тогда как в методе Грама—Шмидта необходимо строить сложные многомерные аналоги алгоритма Горнера. Кроме того, по крайней мере в одномерном случае осуществимо вложение (англ. *telescoping*) ортонормированных аппроксимирующих рядов [38, 43, 48]. Мы предполагаем, что вложение в принципе возможно и для многомерных рядов, но нам не известны соответствующие программные реализации. Это означает, что рекуррентный подход более экономичен и с точки зрения объема основных вычислений.

Количественные данные для сравнения при степенях 1—6 для одной, двух и трех независимых переменных приведены в табл. 1.

Проведенное сравнение имеет принципиальный характер и основывается на внутренних свойствах двух методов ортогонализации. В этой связи следует отметить весьма интересную реализацию [17] метода Грама—Шмидта, в которой использовалась простая мультипликативная связь между полиномами естественного базиса для случая трех переменных, а также был выбран наиболее выгодный порядок вычисления скалярных произведений. На основе полученных результатов автор [17] сделал вывод о том, что метод Грама—Шмидта имеет преимущества по сравнению с рекуррентными методами ортонормировки. Мы не разделяем этих взглядов, так как считаем, что не следует заменять сравнение методов сравнением их конкретных реализаций. Класс быстрых алгоритмов типа быстрого преобразования Фурье применим, в принципе, и к рекуррентным методам ортогонализации. Как видно из табл. 1, относительная разность объемов необходимых расчетов падает с повышением числа независимых

Таблица 1. Сравнение чисел скалярных произведений, чей расчет переменных по методам Грама —

Максимальная степень системы полиномов	Размерность 1		Раз	
	Число линейно-независимых полиномов в системе	Метод		
		Грама — Шмидта	рекурсии Форсайта — Вайсфельда	Число линейно-независимых полиномов в системе
1	2	3	3	3
2	3	6	5	6
3	4	10	7	10
4	5	15	9	15
5	6	21	11	21
6	7	28	13	28

переменных. Тем не менее из-за большого абсолютного объема вычислений следует пользоваться всеми доступными средствами, приводящими к улучшению быстродействия программ,— в том числе более экономным методом в сочетании с удачным алгоритмом.

Дополнительным преимуществом ортогональных разложений (независимо от способа ортогонализации и помимо улучшения обусловленности) является взаимная независимость коэффициентов ортогонального ряда, простота расчета полных погрешностей коэффициентов и самих рядов, а также легкость сравнения рядов различной длины J .

Резюмируя изложенное выше, отметим еще раз три основных преимущества ортонормировки Форсайта—Вайсфельда, полностью окупающие дополнительные вычисления на ее реализацию: 1) быстрый и простой алгоритм; 2) нормировка, обеспечивающая малую погрешность округления (см. также [39]); 3) упрощение вычислений при подборе нужной степени J полиномов, так как используются без изменения коэффициенты ранее вычисленных полиномов [16, 37], см. также [39].

Вообще говоря, скалярное произведение является преобразованием типа свертки и в наиболее общем виде его можно записать как интеграл Лебега [40]:

$$(f, g) = \int_{\Omega} f(x) g(x) d\mu, \quad (28)$$

где f и g интегрируемы на множестве Ω неотрицательной меры μ . Независимая переменная x в данном случае необязательно одномерна. Поскольку у нас x — координатный континуум, то без ограничения общности можно записать

$$d\mu = w(x) dx,$$

необходим для ортогонализации полиномов от одной, двух и трех Шмидта и Форсайта—Вайсфельда

мерность 2		Число линейно-независимых полиномов в системе	Размерность 3	
Метод			Метод	
Грама — Шмидта	рекурсии Форсайта — Вайсфельда		Грама — Шмидта	рекурсии Форсайта — Вайсфельда
6	6	4	10	10
21	19	10	55	52
55	46	20	240	191
120	92	35	630	541
231	162	56	1596	1282
406	261	84	3570	2675

где $w(x)$ — неотрицательная весовая функция. Тогда различные скалярные произведения будут отличаться областью интегрирования Ω и весовой функцией $w(x)$. Видно, что имеется широкий выбор допустимых скалярных произведений. Некоторые из них хорошо изучены и систематизированы: например, для одномерного аргумента и гладких весовых функций существует всего пять принципиально различных семейств ортогональных полиномов [43, 44]: Лежандра, Чебышева, Эрмита, Лагерра и Якоби — так называемые классические ортогональные многочлены.

Если реперные точки более или менее равномерно распределены по полю чувствительности Ω калибруемого прибора, а точность определения координат реперов приблизительно одинакова, то наиболее естественно положить

$$w(x) \equiv 1,$$

затем элементарным линейным преобразованием привести Ω к единичному отрезку (или квадрату при $n = 2$ и кубу при $n = 3$) и использовать скалярное произведение вида

$$(f, g) = \int_{\Omega} f(x) g(x) dx. \quad (29)$$

Ортонормировка базиса при помощи (29) не приводит матрицу G из (25) к диагональному виду, но все же делает ее «квазидиагональной», т. е. с резко преобладающей главной диагональю. При этом обусловленность G_J заметно улучшается по сравнению с «естественным» базисом и численное решение системы (25) оказывается гораздо более устойчивым. Указанный подход был использован в работе [41], где двумерная ортогонализация была проведена аналитически посредством системы REDUCE [45] на основе метода Грама — Шмидта. Эта техника была использована для выражения прямых калибровочных преобразований (т. е. для получения ожидаемых позиций реперов исходя из их известных действительных позиций), поскольку в этом случае как ортогонализация, так и обращение матрицы Грама проводится только один раз, а затем используется многократно. В этом смысле можно считать, что трудоемкость подготовительных расчетов — ортогонализации и обращения — имеет второстепенное значение.

Точность определения позиций реперов $p(x)$, конечно, никогда не является одинаковой по всему полю Ω и во многих случаях ухудшается по краям. Поэтому представляется полезным использование весовой функции вида

$$w(x) = 1/p^2(x), \quad (30)$$

если можно предложить разумную аналитическую модель точности $p(x)$. Веса (30) имеют больший физический смысл и лучше соответствуют характеристикам калибруемого прибора. Пример их практи-

ческого использования в случае спирального измерителя будет дан ниже.

При проводимых периодически прямыми калибровками может оказаться, что набор измеренных координат реперов не постоянен (не на каждом кадре видны все реперы), а сами координаты и точности их определения подвержены дрейфу; тогда было бы желательно применять быстрый и эффективный метод ортонормировки, приводящий к единичной матрице G_J для конкретного набора координат и их ошибок. В этих целях удобнее вернуться к общему определению (28) и ввести меру в виде

$$d\mu = \sum_{i=1}^N w_i \delta(x - x_i) dx, \quad (31)$$

где вес w_i выражается через точность ΔF_i определения i -го репера:

$$w_i = 1/(\Delta F_i)^2,$$

$\delta(x - x_i)$ — дельта-функция Дирака; x_i — наблюдаемые координаты i -го репера, а N — общее число реперов. Подстановка (31) в (28) приводит к выражению скалярного произведения через конечную сумму:

$$(f, g) = \sum_{i=1}^N f(x_i) w_i g(x_i). \quad (32)$$

Базис, полученный при помощи скалярного произведения этого типа, называется ортонормированным на дискретном точечном множестве. Поскольку множество это было выбрано совпадающим с множеством наблюдаемых координат реперов, а веса определены в соответствии с точностью определения координат, то нетрудно убедиться, что в этом базисе

$$G_J = I \quad (33)$$

и уравнение (26) непосредственно дает $\mathbf{a}_J = \mathbf{f}_J$, т. е. коэффициенты разложения искомого преобразования калибровки являются скалярными произведениями наблюдаемого вектора \mathbf{F} на соответствующие базисные векторы. Из (33) видно также, что обусловленность решаемой задачи (25) единична (т. е. оптимальна).

Известно, что при скалярном произведении вида (29) число элементов генерируемого ортонормированного базиса не ограничено. Однако так как ограничено само число реперов N , в принципе можно использовать только первые N полиномов, а на практике — гораздо меньше, чтобы избежать превращения задачи калибровки в задачу интерполирования. В работе [16] предложен способ оценки скалярного квадрата очередного полинома ортонормированного базиса, не прибегающий к расчету значений полинома и основанный на значениях рекуррентных коэффициентов. Если этот квадрат отрицателен или меньше по модулю относительной точности ЭВМ, то это является

указанием, что размерность пространства Q_M достигнута и дальнейшее построение высших в лексикографическом порядке полиномов невозможно.

В некоторых практических случаях простейшие геометрические или физические соображения приводят к выводу о необходимости ограничения степеней по одной (или нескольким) из координат. Это осуществимо наложением добавочных ограничений на максимально допустимые степени по некоторым переменным, как это указано пунктиром на рис. 4. Иной пример построения «неполной» системы

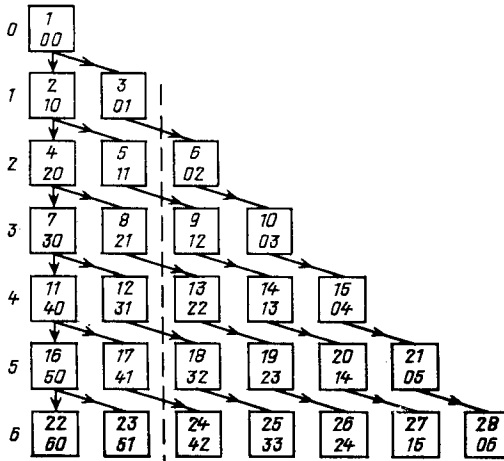


Рис. 4

Рис. 4. Иерархия и лексикографический порядок полиномов от двух переменных степеней 0–6:

справа указана общая степень. Число сверху каждой клетки — порядковый номер полинома; двойка чисел под ним — распределение степеней. Если наложено условие линейности по оси J , необходимо пропустить (или считать тождественными нулями) все члены семейства справа от вертикального пунктира

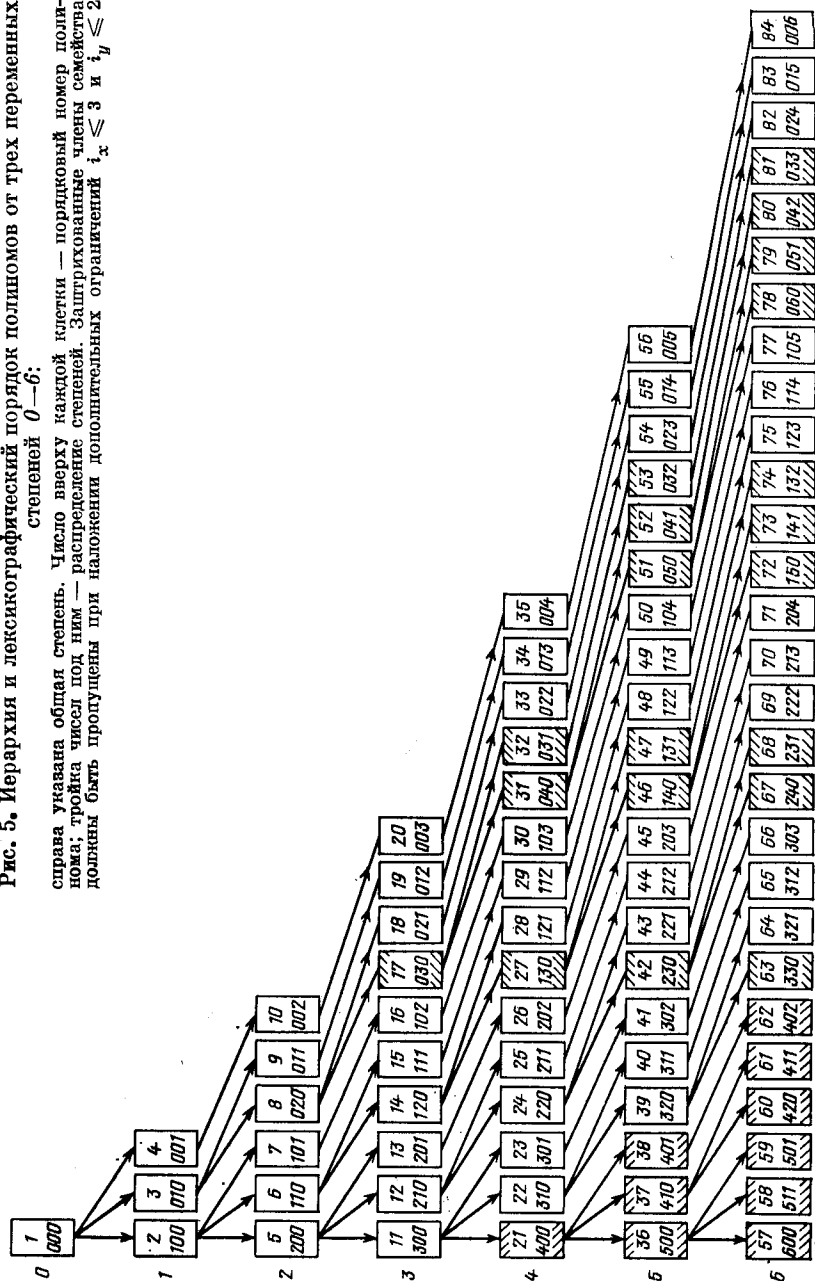
ортонормированных полиномов от трех переменных с ограничениями по степеням $i_x \leq 3$, $i_y \leq 2$ и $i_z \leq 6$ показан на рис. 5, где изображена полная иерархия системы до степени 6 и заштрихованы пропускаемые при указанных ограничениях элементы базиса. У нас нет опыта использования неполных ортонормированных базисов, кроме очевидных одномерных случаев разложения четных функций по четным полиномам и нечетных — по нечетным. Нам известно, что таким образом проводилась аппроксимация магнитного поля камеры РИСК [46], но результаты опубликованы не были.

В предыдущем изложении мы не накладывали никаких условий на пространственное расположение реперных точек в области Ω . Важно отметить, что если Ω представима как прямое произведение своих подпространств

$$\Omega = \Omega_1 \times \Omega_2, \tag{34}$$

Рис. 5. Иерархия и лексикографический порядок полиномов от трех переменных степеней 0 — 6:

справа указана область степеней. Число сверху каждой клетки — порядковый номер полинома; тройка чисел под ним — распределение степеней. Загнтрихованные члены семейства должны быть пропущены при наложении дополнительных ограничений $i_x \leq 3$ и $i_y \leq 2$



то возможно разделение переменных и представление базисных полиномов в Ω в виде произведений элементов базиса из Ω_1 и Ω_2 , причем ортонормировка в Ω_1 никоим образом не связана с ортонормировкой в Ω_2 , и наоборот. Естественно, так как размерности подпространств ниже размерности Ω , такое представление гораздо более экономично в программной реализации — ср. размерности $n = 1$, $n = 2$ и $n = 3$ табл. 1. Сетки реперов, для которых (34) возможно, можно назвать псевдопериодическими: они, подобно периодическим решеткам, состоят из «элементарных» ячеек подобной формы, но разных размеров — рис. 6.

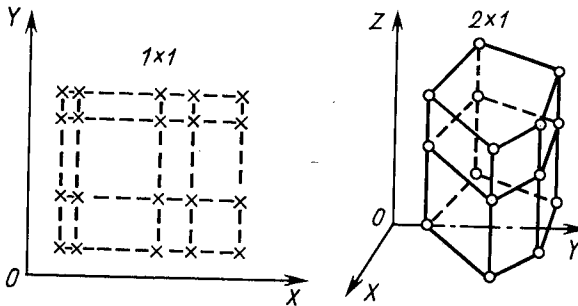


Рис. 6. Примеры решеток, допускающих разделение переменных

И еще одно замечание. Пока мы везде подразумевали декартову прямоугольную систему координат, в которой полиномиальные представления наиболее наглядны. Однако рассматриваемая рекуррентная ортонормировка не требует непременно именно такой координатной системы. Например, не существует никаких принципиальных препятствий проведению ортонормировки непосредственно в полярных координатах, минуя многократные трудоемкие переходы к декартовой системе, особенно если речь идет об изображении локальных объектов (реперных крестов). При работе со всем полем необходимо учесть непрерывность и гладкость аппроксимирующей функции при $\theta = 0$ и $\theta = 2\pi$.

Теперь, предположив, что процесс ортонормировки базиса закончен, рассмотрим собственно аппроксимацию калибровочных зависимостей посредством конечных рядов вида (24) при размерности M линейного пространства Q_M , причем $J \leq M \leq N$. Заметим, что при $J = M$ (N достижимо, только если реперные точки не расположены на алгебраической поверхности низшего порядка) это выражение переходит в известную интерполяционную формулу Лагранжа [40] или же в ее многомерные аналоги [32], что нежелательно хотя бы по двум причинам: а) наши реперы, разумеется, измерены с конечной точностью, и нет никаких оснований делать аппроксимацию точной в соответствующих точках; б) интерполяция высоких порядков по

Лагранжу чрезвычайно нестабильна в междоузлиях. Следовательно, первый вопрос при использовании (24) касается определения оптимального порядка калибровки аппроксимации J . Вообще говоря, метод ортонормированных разложений открывает относительно критериев обрывания ряда [47], а подбор подходящего критерия при калибровке конкретного прибора является задачей более физической, нежели математической. Нам не представляется возможным сформулировать всеобщий формальный критерий, который был бы хорош на все случаи жизни. Поэтому мы ограничимся перечислением наиболее часто используемых критериев:

1. Достижение абсолютного минимума взвешенной суммы остаточных квадратов $\sum_{i=1}^N w_i [F_i - F_J(x_i)]^2$ для $J = 1, 2, \dots, M - 1$ (значение $J = M$ исключается по причинам, указанным выше). Эта величина не всегда монотонна по J из-за накопления погрешностей округления. Мы не рекомендуем ее использование, так как она более характерна для вычислительной среды, чем для решаемой задачи.

2. Достижение минимума χ^2 , нормированного на одну степень свободы:

$$\chi_{N-J}^2 = \sum_{i=1}^N w_i [F_i - F_J(x_i)]^2 / (N - J) \rightarrow \min. \quad (35)$$

Это — наиболее часто используемый статистический критерий [35, 48]. Он корректен, но опираться только на него небезопасно.

3. Критерий Фишера [49], т. е. отсутствие значимого отличия от нуля двух высших коэффициентов разложения, следующих за J .

4. Выполнение условия [21, 38, 48]:

$$w_i [F_i - F_J(x_i)]^2 \leq 1, \quad i = 1, 2, \dots, N. \quad (36)$$

Геометрический смысл этого критерия состоит в прохождении аппроксимирующей кривой (поверхности) через коридоры ошибок всех реперных точек. Критерий очень надежен при несмещенных оценках позиций реперов и справедливых оценках ошибок (т. е. весов).

5. Достижение априорной абсолютной $|F_i - F(x_i)| \leq \Delta F$ или относительной $|(F_i - F(x_i))/F_i| \leq \Delta f$ точности во всех реперных точках. Этот критерий подходит для быстрой текущей (проверочной) калибровки, если точностные характеристики прибора Δf или ΔF были определены заранее иными способами.

6. Критерий минимакса или равномерной аппроксимации

$$\min_J \max_i |F_i - F_J(x)|.$$

7. Критерии гладкости, например:

$$\min_{J'} \max_i \left| \frac{dF_{J'}(x_i)}{dx} \right|,$$

где диапазон J' сужен по сравнению с J во избежание очевидных случаев $\frac{dF}{dx} \Big|_{J=1} \equiv 0$ и $\frac{dF}{dx} \Big|_{J=2, 3, \dots, n+1} = \text{const}$ или более трудоемкий, но и более эффективный критерий монотонности аппроксимирующего ряда в междоузлиях [50].

8. Различного вида комбинированные критерии, например, минимум произведения трех факторов

$$\chi_{N-J}^2 \min_J \max_i |F_i - F_J(x_i)| \min_J \max_i \left| \frac{F_i - F_J(x_i)}{F_i} \right|,$$

использованный в [51, 52] для решения задачи, не имеющей отношения к калибровке, но тем не менее аппроксимационной.

Приведенный список не претендует на полноту. Он просто иллюстрирует богатую гамму известных критериев и еще раз подчеркивает необходимость проникновения в физическую суть задачи для выбора наиболее подходящего из них.

Расчет коэффициентов разложения $\{a_j\}$ таблично заданной функции $\{F_i\}$ по ортонормированному базису $\{b_j\}$ не требует обращения матриц и сводится к вычислению скалярных произведений

$$a_j = (F, b_j). \quad (37)$$

Как известно, набор коэффициентов $\{a_j\}$ взаимно независим в любом ортогональном базисе. В работе [53], где используется ортогональный, но не нормированный базис, справедливо указывается, что применение эквивалентной формулы

$$a_j = (F_j, b_j), \quad (38)$$

где

$$F_j = F - \sum_{h=1}^{j-1} a_h b_h,$$

приводит к меньшему эффекту накопления погрешностей округления. Это действительно так, хотя и проявляется менее четко при нормировке базиса. Формула (38) требует большего объема элементарных вычислений, чем (37), и, кроме этого, при ее использовании независимость между коэффициентами теряется. К этому методу можно было бы прибегнуть, когда доподлинно известно, что именно накопление погрешностей округления является причиной неудовлетворительного качества аппроксимации. В нашей практической работе по калибровке различных трековых измерительных систем ортонормированными полиномами мы не встречали подобных ситуаций. Всегда можно было использовать более простую формулу (37) и сохранить независимость коэффициентов разложения.

Полные среднеквадратичные отклонения $\{\Delta a_j\}$ не зависят от индекса j и численно равны [35, 49]:

$$\Delta a_j = \sqrt{\chi_{N-J}^2}, \quad j = \overline{1, J}. \quad (39)$$

Кстати, величину χ_{N-J}^2 можно выразить, не прибегая к суммированию ряда (24), но этим свойством не следует пользоваться из-за опасности потери точности при вычитании близких чисел в представлении с плавающей запятой.

Из (24) и (39) при независимых коэффициентах $\{a_j\}$ непосредственно следует, что точность аппроксимирующей функции ΔF в любой точке x выражается как

$$\Delta F(x) = [\chi_{N-J}^2 \sum_{j=1}^J b_j^2(x)]^{1/2}.$$

Опять-таки, ортонормировка базиса $\{b_j\}$ обеспечивает простой и надежный способ оценки $\Delta F(x)$.

В заключение данного раздела можно отметить, что к настоящему времени преимущества ортонормированных полиномиальных базисов не вызывают сомнений. Их аппарат достаточно разработан теоретически, технические детали его алгоритмизации выяснены, а его программная реализация не представляет серьезных трудностей. Имеющийся опыт эксплуатации программ полностью подтверждает эти выводы.

3. ОПРЕДЕЛЕНИЕ ЦЕНТРОВ КРЕСТОВ

Первый этап статистической обработки данных калибровочных измерений состоит в распознавании крестов, нахождении центров и погрешностей в их определении. Эта задача важна для обработки не только калибровочных измерений, так как реперный крест — это одна из самых массовых фигур на снимках с трековых камер. Центры реперов необходимы для восстановления пространственной картины события и для учета возможных дисторсий прибора. Поэтому центры подлежат особо точному измерению. В этой связи необходимо учитывать нелинейность измерительных устройств, особенно существенную для приборов с электронно-лучевым сканированием и в системах бесфильмового съема данных, где возможны значительные искажения прямолинейности плеч некоторых крестов и угла между ними [54]. Тем не менее практически во всех случаях наличие искажений не препятствует выделению прямоугольных областей примерного расположения каждого из крестов на основе априорных сведений о самой калибровочной решетке. Таким образом, процедура распознавания и выделения центров крестов обычно осуществляется в цикле по всем крестам решетки, где для каждого креста выполняются следующие пять операций:

а) выделение области расположения всех точек креста и перенесение этой области в начало координат для упрощения дальнейших вычислений;

б) разбиение всех точек области на классы, относящиеся к двум плечам креста, и исключение шумовых точек;

- в) аппроксимация каждого из плеч креста;
- г) определение координат центра креста как совместное решение уравнений плеч;

д) вычисление погрешностей координат центров.

Эта схема в основном сохраняется и в случае спирального сканирования, хотя там есть своя специфика, накладываемая двумя координатными системами.

Распознавание креста. Операция выделения области, включающей все точки k -го креста для симметричной решетки, фактически уже была описана в начале разд. 2. Априорные оценки масштабов по осям M_{x_0} , M_{y_0} , а также коэффициентов переноса \hat{X}_0 , \hat{Y}_0 между системами координат прибора и идеальной системой, связанной с изображением, позволяют определить преобразование переноса центра этой области в начало координат:

$$u'_{ik} = u_{ik} - U_{k0}; v'_{ik} = v_{ik} - V_{k0}; i = \overline{1, N},$$

где U_{k0} , V_{k0} определены в (12), i — индекс текущей точки.

Переход в подобную локальную систему декартовых координат выглядит много сложнее в случае обработки данных спирального сканирования [21], где границы

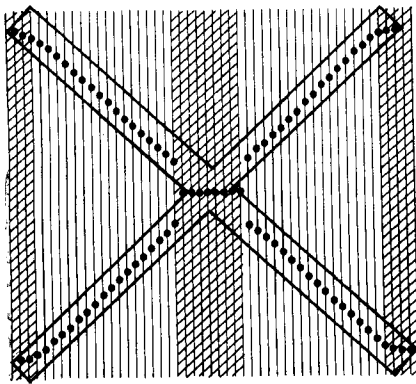


Рис. 7. Пример искажений оцифровки креста

области в полярной системе координат R , θ образуют симметричную трапецию, а переход в локальную декартову систему, даже приближительный, требует, кроме оценок масштабов M_R и M_θ по осям, предварительной оценки θ_0 начального угла отсчета. Кресты калибровочной решетки СИ в локальной системе разворачиваются так, чтобы ось OX' шла приблизительно вдоль биссектрисы меньшего угла креста с тем, чтобы его плечи удовлетворяли уравнению (13).

Следует сразу же указать, что в любой системе, выпол-

няющей глобальное сканирование всей калибровочной решетки, линия сканирования идет под углом к плечам креста. Конечная разрешающая способность сканирующего пятна приводит в этом случае к смещению отсчетов на концах крестов и к их слиянию в середине креста (рис. 7). Точки из этих участков, заштрихованных на рис. 7, следует устранить, т. е. попросту не переводить в локальную систему координат.

Выбор метода распознавания креста зависит от искажений, вносимых измерительной системой, и зашумленности изображения.

Когда и то и другое мало, можно воспользоваться одним из методов кластеризации (см., например, [26]), т. е. объединения точек по некоторым критериям близости. Учет близости по расстоянию ρ_i

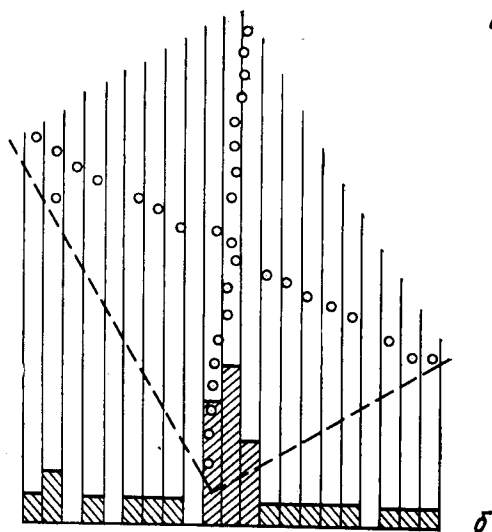
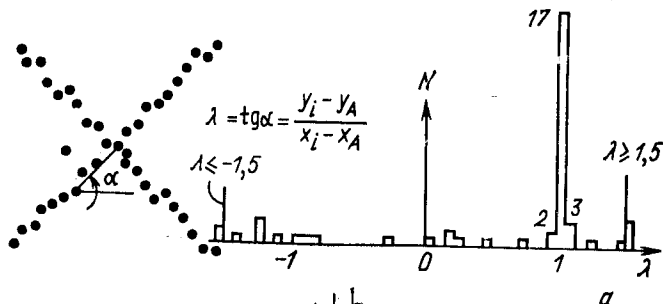


Рис. 8. Способы распознавания креста:
 а — метод кластеризации, б — метод гистограммирования

направлению (с помощью проверки близости угловых коэффициентов прямых, соединяющих выбранную точку со всеми остальными, рис. 8, а) позволяет распознать оба плеча креста, даже если неизвестна их ориентировка. Однако подобная «либеральность» этого метода делает его малоустойчивым к шумам вообще и особенно к засорениям в виде посторонних штрихов, царапин и т. д.

Более помехоустойчивым и потому наиболее употребительным является метод гистограммирования точек по заданным направлениям с выбором интервала, дающего максимальный вклад в гистограмму (рис. 8, б). За счет объединения интервалов, соседних с ма-

ксимальным, этот метод находит плечи креста, даже если направление задано приблизительно (рис. 8, б).

В случаях, когда дисторсии измерительных устройств настолько малы, что не меняют угла раствора креста, наиболее эффективен метод определения креста как целого [55], т. е. как вырожденной кривой второго порядка.

Перемножив уравнения плеч (13), исключим как малые второго порядка члены, содержащие произведения малых по предположению величин Δ_k , Δ_{1k} , Δ_{2k} , и введем новые обозначения $a = \Delta_{1k} + \Delta_{2k}$, $b = -\lambda(\Delta_{2k} - \Delta_{1k})$, $c = 2\Delta_k$.

Получим функционал

$$\Phi_k = \sum_{i=1}^N (v_{ik}^2 - \lambda^2 u_{ik}^2 - av_{ik} - bu_{ik} - cu_{ik}v_{ik})^2,$$

нормальная система уравнений которого линейна относительно новых переменных. Из-за приближенности функционала точное решение требует нескольких повторных итераций. Однако в [21] проводит-

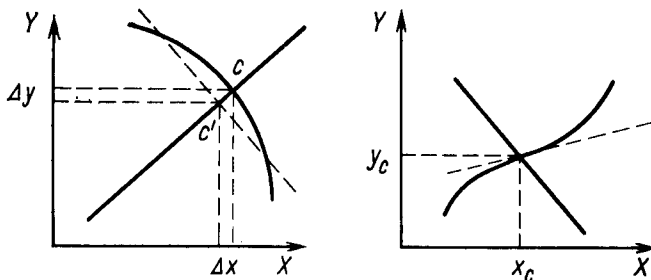


Рис. 9. Характер искажений при неучитывании локальной нелинейности: а — сдвиг центра; б — завышение погрешностей

ся только первая итерация. Полученной точности в определении Δ_k , Δ_{1k} , Δ_{2k} по найденным a , b , c оказывается достаточно для классификации всех точек из области креста на две группы в соответствии в их принадлежностью к плечам креста.

Предположения о малости дисторсий выполняются далеко не всегда. Например, предварительные эксперименты по калибровке бесфильмовой системы съема информации показали, что даже в областях отдельных крестов отчетливо наблюдаются нелинейные искажения плеч (рис. 9). Причины нелинейности здесь связаны с нелинейностью развертки электронного луча, с дисторсией усилителей яркости света (ЭОП), с нестабильностью электронных схем сканирования, с абберациями оптических систем [10]. Поэтому плечи реперных крестов нельзя во всех случаях считать отрезками прямых [54], а угол между ними — совпадающим с соответствующим углом для калибровочной решетки.

В этой связи в качестве наиболее универсального метода, не требующего проверки малости дисторсий, предлагается метод одновременной робастной подгонки обоих плеч креста с помощью полиномов подходящих степеней.

Определение центров крестов. Если обозначить координаты точки пересечения двух плеч через u_c, v_c , то можно записать уравнения плеч в виде переопределенной системы:

$$\left. \begin{aligned} f_1(u_{1i}, v_{1i}) &= f_1(u_c, v_c), & i &= \overline{1, N_1}; \\ f_2(u_{2i}, v_{2i}) &= f_2(u_c, v_c), & i &= \overline{1, N_2} \end{aligned} \right\} \quad (40)$$

и искать неизвестные u_c, v_c как координаты точки, минимизирующие сумму остаточных квадратов уравнений, т. е. функционал

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{N_1} w_i [f_1(u_{1i}, v_{1i}) - f_1(u_c, v_c)]^2 + \\ &+ \sum_{i=1}^{N_2} w_i [f_2(u_{2i}, v_{2i}) - f_2(u_c, v_c)]^2. \end{aligned} \quad (41)$$

Важен вопрос о выборе весов, обеспечивающих устойчивость решения к засоряющим данным. Специфика задачи (хорошее начальное приближение для направлений плеч и отсутствие ситуаций типа той, что представлены на рис. 3) позволяет использовать в качестве весовой функции не (22), а более простой «бивес» Тьюки [30]:

$$w_i = (1 - (\hat{d}_i / 5\hat{\sigma})^2)^2$$

с $\hat{\sigma}$, оцененной по медиане $|\hat{d}_i|$ для начальных значений параметров.

Прежде чем перейти к описанию метода решения поставленной задачи, необходимо рассмотреть вопрос о типах нелинейностей функций f_1 и f_2 , что позволит наметить качественный путь ее решения.

Наблюдаются следующие типы нелинейностей (рис. 9): а) практически неискаженные плечи, которые достаточно точно аппроксимируются прямыми; б) выпуклые или вогнутые плечи с кривизной постоянного знака. Эту форму можно с достаточной точностью представить в виде полинома второй степени; в) S-образные плечи, у которых кривизна меняет знак. Хорошим приближением к ним служит полином третьей степени.

Тогда f_1 и f_2 могут быть представлены в виде

$$f_j = \sum_{i=0}^{l_j} c_{ji} u_j^i - v_j, \quad j = 1, 2, \quad (42)$$

где $l_j = 1, 2, 3$ и каждое из них можно определить индивидуально и независимо для каждого плеча. Кроме этого, искажения могут вызывать изменение углов между плечами креста, поэтому мы отказы-

ваемся от использования предположения о перпендикулярности плеч и допускаем только, что плечи реперного креста пересекаются под определенным, но неизвестным углом. Это допущение освобождает нас от необходимости перехода к линейным шкалам на раннем этапе калибровки и, помимо этого, упрощает алгоритм (отпадает необходимость в построении касательных к кривым f_1 и f_2 в точке (u_c, v_c)) и делает его более универсальным, так как позволяет использовать косоугольную реперную сетку и различные масштабы по осям u и v .

Подстановка уравнений (42) в (40) и минимизация функционала (41) приводит к нелинейной системе нормальных уравнений, которую можно разложить на две независимых линейных системы типа (17) относительно $\{c_{ji}\}$; детальное описание численно-аналитического алгоритма ее решения можно найти в [54].

При бесфильмовом методе необходимо учитывать дополнительные искажения, возникающие при сжатии информации, т. е. замены плеч крестов линейными треугольными элементами [11, 56]. В [54] проведено также сравнение точности найденных координат центра крестов при различных методах предварительной обработки информации (например, со сжатием данных и без него).

Наиболее радикальным примером применения метода робастной полиномиальной подгонки плеч креста может служить идея использования этого метода для непосредственного определения полярных координат центров крестов при спиральном сканировании, минуя трудоемкие операции перевода всех точек в локальную систему декартовых координат и обратно.

Вычисление погрешностей координат центров. При совместном решении аппроксимирующих уравнений плеч креста координаты точки их пересечения, т. е. центра (u_c, v_c) , определяются как функция от коэффициентов аппроксимирующих полиномов. Для каждого из этих полиномов в ходе стандартной МНК-процедуры вычисляются ковариационные матрицы коэффициентов, что позволяет рассчитать и погрешности координат u_c, v_c . Для этого необходимо подставить найденные значения u_c, v_c в функционал (41) и рассчитать их погрешности, составив якобиан уравнений в точке решения.

В качестве примера более простого пути рассмотрим случай $l_1 = l_2 = 1$. В соответствии с (18) получаем оценки коэффициентов \hat{a}_j, \hat{b}_j ; $j = 1, 2$, определяющие две прямые $v = \hat{a}_j u + \hat{b}_j$, и находим координаты точки их пересечения:

$$\hat{u}_c = (\hat{b}_2 - \hat{b}_1) / (\hat{a}_1 - \hat{a}_2); \quad \hat{v}_c = (\hat{a}_1 \hat{b}_2 - \hat{a}_2 \hat{b}_1) / (\hat{a}_1 - \hat{a}_2). \quad (43)$$

Следуя [57], разлагаем \hat{u}_c и \hat{v}_c в ряд Тейлора относительно неизвестных средних значений параметров, т. е. a_j, b_j ; $j = 1, 2$. Ограничиваясь только первым членом разложения и усредняя квадраты (\hat{u}_c —

$-u_c)^2$ и $(v_c - v_c)^2$, получаем приближенные формулы для дисперсий:

$$\sigma_{u_c}^2 \approx [(\sigma_{a_1}^2 + \sigma_{a_2}^2) \hat{u}_c^2 + 2(\rho_{a_1 b_1} \hat{a}_1 + \rho_{a_2 b_2} \hat{a}_2) \frac{\hat{u}_c}{(\hat{a}_1 - \hat{a}_2)} + (\sigma_{b_1}^2 + \sigma_{b_2}^2)] / (\hat{a}_1 - \hat{a}_2)^2;$$

$$\sigma_{v_c}^2 \approx [(\hat{a}_2^2 \sigma_{a_1}^2 + \hat{a}_1^2 \sigma_{a_2}^2) \hat{u}_c^2 +$$

$$+ 2(\hat{a}_2^2 \rho_{a_1 b_1} \hat{a}_1 + \hat{a}_1^2 \rho_{a_2 b_2} \hat{a}_2) \frac{\hat{u}_c}{(\hat{a}_1 - \hat{a}_2)} + (\hat{a}_2^2 \sigma_{b_1}^2 + \hat{a}_1^2 \sigma_{b_2}^2)] / (\hat{a}_1 - \hat{a}_2)^2,$$

где величины $\sigma_{a_j}^2$, $\sigma_{b_j}^2$, $\rho_{a_j b_j}$; $j = 1, 2$, вычисляются по формулам (19).

Как уже отмечалось, в случае спирального сканирования при использовании перевода данных каждого креста в локальную систему координат для получения там координат центра (43) происходит поворот на угол $\theta_k = \arctg(v_k/u_k)$, что приводит к появлению в матрице ошибок отличных от нуля корреляционных членов вида $\sigma^2 (\lambda^2 - 1) \cos \theta_k \sin \theta_k$ [21]. Это значение, например при $\theta_k \sim \pi/4$, уже сравнимо со значением диагональных членов матрицы ошибок и должно учитываться при оценках калибровочных параметров.

4. ПОСТРОЕНИЕ КАЛИБРОВОЧНЫХ ПРЕОБРАЗОВАНИЙ

Приступая ко второму этапу калибровки, мы располагаем наборами пар идеальных $\{x_i, y_i\}_{i=1}^N$ и измеренных $\{u_{jk}, v_{jk}\}_{j=1}^{N'_k} \underset{k=1}{L}$ координат, где $N'_k \leq N$ (по тем или иным причинам в k -м измерении могли быть пропущены некоторые реперы), а L — число измерений. В зависимости от способа получения набора $\{u_{jk}, v_{jk}\}$ в полуавтоматическом или автоматическом режиме сканирования решетки он может быть либо полным, как это указано выше, либо же представлять собой значения, усредненные по измерениям: $\{u_j, v_j\}_{i=1}^{N'_j}$. Вообще говоря, усреднение может быть проведено параллельно с построением калибровочных преобразований и не обязано предшествовать им. В принципе, усреднение допустимо только для однородных данных. При нарушении условия однородности (например, сдвиг начала или направления полярной оси в СИ) приходится усреднять после учета или компенсации неоднородностей. Во всех случаях существенным элементом данных являются также точности измерения $\{\Delta u_{jk}, \Delta v_{jk}\}$ или $\{\Delta u_j, \Delta v_j\}$. В дальнейшем мы будем пользоваться обозначениями с одним индексом, подразумевая обе указанные ситуации.

Прямое и обратное преобразования. Общепринято называть прямым преобразованием калибровки операцию $\{u_j, v_j\} \xrightarrow{F} \{x_i, y_j\}$, переводящую измеренные координаты в идеальные, и обратным преобразованием — соответственно $\{x_i, y_i\} \xrightarrow{F^{-1}} \{u_i, v_i\}$. Хотя термины эти и укоренились в практике, в них не следует вкладывать ма-

тематический смысл, с точки зрения которого каждая из задач нахождения явного вида как прямого, так и обратного калибровочных преобразований является типичной обратной задачей со всеми присутствующими таким задачам особенностями: некорректность, необходимость привлечения априорной информации и т.п. [58].

Явный вид операторов F и F^{-1} находят приближенно, пользуясь линейной аппроксимационной моделью (24). Можно указать четыре типа используемых при этом базисов $\{\varphi_i\}$ [14, 59]:

$\{\varphi_i^{(1)}\}$ — естественный базис $1, x, y, x^2, xy, y^2 \dots$ (см. разд. 2);

$\{\varphi_i^{(2)}\}$ — базис, ортогонализированный по [45] на единичном квадрате $x, y \in [-1, 1]$ со скалярным произведением вида (29);

$\{\varphi_i^{(3)}\}$ — ортогональный на дискретном фиксированном множестве базис, получаемый из набора идеальных координат методом (26);

$\{\varphi_i^{(4)}\}$ — базис, ортонормированный на произвольном дискретном множестве по рекуррентному методу Форсайта — Вайсфельда с произвольными положительными весами.

Отметим, что $\{\varphi_i^{(1)}\}$ включен из-за своей наглядности и для логической полноты, тогда как его практическое значение ограничено случаями отсутствия дисторсий. Обусловленность в $\{\varphi_i^{(2)}\}$ и $\{\varphi_i^{(3)}\}$ намного лучше; они применимы соответственно к прямому и обратному преобразованиям. Базис $\{\varphi_i^{(4)}\}$ наиболее универсален и им можно пользоваться в обоих случаях с оптимальной (единичной) обусловленностью. Сравнение наборов $\{\varphi_i^{(1)}\} \div \{\varphi_i^{(4)}\}$ можно найти в [60].

При расчете явного вида F обычно пользуются весами

$$w_i = 1 / (\Delta u_i^2 + \Delta v_i^2),$$

рассматривая сам оператор как двухкомпонентный вектор (F_x, F_y) . Обе компоненты разлагаются по тому же базису, но с разными коэффициентами — см. (8). Поскольку указанные веса выражают степень соответствия i -й измеренной пары идеальной, ими же следует пользоваться и для численного построения базиса обратного преобразования F^{-1} . Конечно, из-за нелинейности F относительно координат, а также по причине его приближенности прямое обращение $F \rightarrow F^{-1}$ невозможно.

Вопрос об оптимальной длине J аппроксимирующего ряда (24) решают при помощи подходящего критерия (или совокупности критериев — см. разд. 2), подобранного с учетом особенностей калибруемого прибора. Например, $J = 3$ для просмотрово-измерительного стола САМЕТ, но $J = 21$ при максимальной степени 5 в случае сканирующих автоматов ERASME и АЭЛТ-2/160. Соответствующее исследование проведено в [14].

Случай спирального сканирования. Поскольку спиральный измеритель использует две системы координат: декартовую $(X, Y)_{СИ}$ и полярную $(R, \theta)_{СИ}$, задача калибровки СИ усложняется. Помимо прямого и обратного преобразований $(X, Y)_{СИ} \leftrightarrow (X, Y)_{КР}$ между

декартовыми координатами СИ и калибровочной решетки, требуется также установить взаимные преобразования $(X, Y)_{\text{КР}} \leftrightarrow (R, \theta_{\text{СИ}})$ и $(X, Y)_{\text{СИ}} \leftrightarrow (R, \theta_{\text{СИ}})$.

В связи с тем что последнее преобразование может быть получено последовательным применением первых двух, его можно не определять.

Исследование точности системы $(X, Y)_{\text{СИ}}$ [61] показало, что можно ограничиться измерением только пяти крестов, расположенных в пересечении и на концах координатных осей $(X, Y)_{\text{КР}}$, чтобы получить шесть коэффициентов аффинного преобразования (2) $(X, Y)_{\text{СИ}} \leftrightarrow (X, Y)_{\text{КР}}$ и вычислить параметры (4) и (5). Косоугольность системы СИ ОИЯИ оказалась отличной от нуля, но малой: $|\kappa| \leq 0,0003$ [21].

Обработка данных $(R, \theta)_{\text{СИ}}$ показала наличие значимых дисторсий, заметных особенно отчетливо на данных сканирования эталонной прямой линии, образующих S-образную кривую с прогибом до 20 мкм. Как было установлено в [61], основная часть этих нелинейных искажений компенсируется простой линейной коррекцией вида

$$\theta = \theta_0 + \alpha R. \quad (44)$$

Однако оставшиеся искажения оказались значимыми (в частности, содержали ложную кривизну, вносящую систематический сдвиг значений импульсов частиц). Исследование формы оставшихся искажений показало [21] наличие разрыва в полюсе системы и характерных изгибов в области от -10 до $+10$ мм, т. е. как раз там, где калибровочная решетка (см. рис. 1, з) не содержит достаточно информации для интерпретации этих искажений. Поэтому результаты использования данных только от калибровочной решетки при самой изощренной обработке (в том числе и с использованием систем полиномов $\{\varphi_i^{(3)}\}$) не могли обеспечить удаление искажений. В этой связи в [21] было предложено осуществлять совместную обработку данных сканирования как калибровочной решетки, так и эталонной прямой линии. Вместо прямой более подходящим мог бы явиться объект в виде многолучевой звезды из шести — восьми пересекающихся отрезков прямых. Однако показанная в [61] изотропность зависимости (44) от α позволила ограничиться сканированием только одной прямой.

С обработки этих данных и начинается определение нелинейных искажений путем вычисления коэффициентов угловой коррекции вида

$$\theta_{\text{корр}} = \theta_0 + \alpha R + \beta/R + \gamma f(R), \quad (45)$$

где $f(R)$ зависит от формы остаточных искажений. В частности, для СИ ОИЯИ $f(R) = R/[1 + c(R - R_M)^2]$.

Отметим, это попытка более сложной коррекции с помощью ортогональных полиномов [61] оказалась неудачной, так как потребовала использования полиномов 13-й степени (исходя из критерия

1 — см. разд. 2), коэффициенты которых резко менялись при случайных малых отклонениях данных.

После определения параметров (45) все данные сканирования калибровочной решетки подвергались коррекции с помощью β и γ , что позволяло использовать для учета оставшихся искажений в искомом преобразовании $(R, \theta)_{\text{СИ}} \rightarrow (X, Y)_{\text{КР}}$ простое шестипараметрическое выражение:

$$\left. \begin{aligned} \hat{x}_{\text{КР}} &= x_0 + M_R (R - R_0) \cos(\theta - \theta_0 - \alpha R); \\ \hat{y}_{\text{КР}} &= y_0 + M_R (R - R_0) \sin(\theta - \theta_0 - \alpha R). \end{aligned} \right\} \quad (46)$$

Нелинейность (46) по параметрам потребовала для минимизации соответствующего функционала привлечения программы FUMILI [62], выполняющей метод линеаризации функционала, но без учета корреляции исходных данных, на которую было указано в конце разд. 3.

Вместо этого можно линеаризовать саму зависимость (46). Раскроем выражения для \cos и \sin тройных углов и, учитывая малость α и R_0 , заменим $\cos \alpha R \sim 1$, $\sin \alpha R \sim \alpha R$, а члены с произведением αR_0 опустим. Мы приходим к выражениям

$$\hat{x}_{\text{КР}} = \sum_{j=1}^8 a_j \varphi_j(R, \theta); \quad \hat{y}_{\text{КР}} = \sum_{j=1}^8 a_j \psi_j(R, \theta), \quad (47)$$

где связь новых и старых параметров, а также вид функций φ_j и ψ_j заданы в табл. 2.

Таблица 2. Коэффициенты калибровочных преобразований

j	a_j	φ_j	ψ_j
1	x_0	1	0
2	y_0	0	1
3	$M_R \cos \theta_0$	$R \cos \theta$	$R \sin \theta$
4	$M_R \sin \theta_0$	$R \sin \theta$	$-R \cos \theta$
5	$M_R R_0 \cos \theta_0$	$-\cos \theta$	$-\sin \theta$
6	$M_R R_0 \sin \theta_0$	$-\sin \theta$	$\cos \theta$
7	$\alpha M_R \cos \theta_0$	$R^2 \sin \theta$	$R^2 \cos \theta$
8	$\alpha M_R \sin \theta_0$	$-R^2 \cos \theta$	$-R^2 \sin \theta$

Нахождение параметров линейных зависимостей (47) и их погрешностей выполняется с помощью минимизации функционала

$$\chi^2 = \sum_{i=1}^N [w_{x_i} (x_i - \hat{x}_i)^2 + 2w_{y_i} (x_i - \hat{x}_i) (y_i - \hat{y}_i) + w_{y_i} (y_i - \hat{y}_i)^2], \quad (48)$$

в котором (x_i, y_i) — идеальные координаты центра i -го креста калибровочной решетки; \hat{x}_i и \hat{y}_i получены подстановкой в (47) измеренных значений R_i , θ_i этих центров, а w_{xi} , w_{yi} , w_{xyi} — обращением полной матрицы ошибок измеренных центров.

Минимизация (48) выполняется с помощью любой стандартной программы взвешенного МНК.

Старые параметры, содержащие информацию об отсчетных системах СИ, могут быть определены по полученным оценкам с помощью простых формул:

$$x_0 = \hat{a}_1; \quad y_0 = \hat{a}_2; \quad M_R = \sqrt{\hat{a}_3^2 + \hat{a}_4^2}; \quad R_0 = \sqrt{\hat{a}_5^2 + \hat{a}_6^2} / M_R;$$

$$\theta_0 = \arctg(\hat{a}_4 / \hat{a}_3); \quad \alpha = \sqrt{\hat{a}_7^2 + \hat{a}_8^2} / M_R$$

Помимо этого, восстановление «старых» параметров необходимо и для решения проблемы обратного калибровочного преобразования, так как (46), несмотря на свою нелинейность, допускает простое обращение:

$$R = R_0 + 1/M_R \sqrt{(x_{\text{КР}} - x_0)^2 + (y_{\text{КР}} - y_0)^2},$$

$$\theta = \theta_0 + \alpha R + \arctg [(y_{\text{КР}} - y_0) / (x_{\text{КР}} - x_0)]$$

Практика показала [21] важность повторного использования данных сканирования прямой линии. Применение к ним преобразования (46) или (47) с коэффициентами, вычисленными по данным сканирования калибровочной решетки, является очень чувствительным критерием правильности параметризации, выбора начальных приближений при минимизации, корректности вычислений.

Исследование и применение карты остатков. Наиболее полным, т. е. действующим на всем поле измерения, показателем качества калибровочного преобразования является карта остатков — величин $dx_i = x_i - \hat{x}_i$, $dy_i = y_i - \hat{y}_i$; $i = \overline{1, N}$ (рис. 2). В этой связи рекомендуется внимательное изучение карты остатков путем ее визуализации, построения гистограмм самих величин dx_i , dy_i , а также модулей $d_i = \sqrt{dx_i^2 + dy_i^2}$, их моментов \bar{d} , σ_d и $\max_i d_i$. При большом числе крестов используют решающие функции, выдающие диагностическое сообщение при выходе любой из величин d_i , \bar{d} , σ_d , $\max_i d_i$ за заданные пороги. Исследование по выбору наилучшей системы полиномов и оптимальной их степени, проведенное в [14], было основано на изучении гистограмм d_i , а также $\max_i d_i$, \bar{d} и σ_d , вычисленных как в узлах калибровочной решетки, так и в серединах между ними. Как показало это исследование, появление значимых остатков в междоузлиях калибровочной решетки может свидетельствовать о том, что она оказалась слишком грубой для описания структуры дистор-

сий. Увеличение степени аппроксимирующего двумерного полинома в таких случаях не помогает, а может даже повредить делу.

Выходом в таких ситуациях является использование информации, заключенной в самой карте остатков. Если проверена стабильность прибора, позволяющая усреднить карту остатков за несколько циклов калибровки для исключения случайных выбросов, то поправки dx и dy , полученные по такой усредненной карте остатков с помощью двумерной интерполяции, могут компенсировать наличие даже значимых (до 10—15 мкм) систематических отклонений. Возможный способ такой интерполяции предлагается в [21].

Для калибровочной решетки из $m \times n$ крестов с расстояниями между ними h_x по горизонтали и h_y по вертикали карта остатков должна быть преобразована в двумерный массив dx_{ij} , dy_{ij} ; $i = \overline{1, m}$, $j = \overline{1, n}$. Поправки dx и dy в произвольной точке (x, y) вычисляются по следующему алгоритму: 1) вычисляются индексы I и J нижнего левого узла той ячейки, в которую попала точка (x, y) :

$$I = [(m + 1)/2 - x/h_x], J = [(n + 1)/2 - y/h_y]$$

([.] — символ целой части); 2) вычисляются величины

$$X = |x - h_x ((m + 1)/2 - I)|, Y = |y - h_y ((n + 1)/2 - J)|$$

и весовые функции в четырех узлах, окружающих точку:

$$\left. \begin{aligned} w_1 &= (h_x - X)(h_x - Y)/h_x h_y; & w_2 &= X(h_y - Y)/h_x h_y; \\ w_3 &= XY/h_x h_y; & w_4 &= (h_x - X)N/h_x h_y; \end{aligned} \right\} \quad (49)$$

3) вычисляются поправки

$$\begin{aligned} dx &= w_1 dx_{IJ} + w_2 dx_{I+1, J} + w_3 dx_{I+1, J+1} + w_4 dx_{I, J+1}; \\ dy &= w_1 dy_{IJ} + w_2 dy_{I+1, J} + w_3 dy_{I+1, J+1} + w_4 dy_{I, J+1}. \end{aligned}$$

При выходе индексов I и J за диапазоны их изменения $\overline{1, m}$ и $\overline{1, n}$ соответственно $dx = dy = 0$. Система весовых функций (49) удобна тем, что на ребрах ячеек оказываются отличными от нуля веса только тех двух узлов, которые ограничивают ребро, а при попадании точки (x, y) в узел решетки (I, J) автоматически получаем $w_1 = 1$, $w_2 = w_3 = w_4 = 0$.

С помощью подобной процедуры двумерной интерполяции более высокого порядка может быть скомпенсирован рост карты остатков, вызванный недостаточной степенью аппроксимирующих полиномов. Выбор разумного компромисса между этими двумя способами компенсации искажений должен определяться в каждом конкретном случае с учетом их быстродействия и того, какой из способов более адекватно осуществляет эту компенсацию.

5. ТОЧНОСТЬ И СТАБИЛЬНОСТЬ

Чтобы карта остатков стала достоверной характеристикой точности калибровочного преобразования, следует учесть статистический вклад каждого креста в минимизируемый функционал, т. е. домножить остатки на значения их весов. Значимые отклонения взвешенных остатков свидетельствуют о несправильной параметризации, что в случае полиномиальной аппроксимации означает неадекватность используемой системы полиномов.

Таким образом, мы приходим к формуле (36), т. е. критерию 4 из разд. 2. Не менее важным является и критерий 2, дающий общую для всех крестов характеристику точности измерений (35). Кроме этого, величина $\min \chi_{N-J}^2$ дает оценку σ^2 -дисперсии на единицу веса.

Поскольку σ^2 определяет матрицу ошибок и обратную ей матрицу весов W , используемую для вычисления и минимизации функционала (14) до того, как станет известна оценка (35), принято либо задавать σ заранее на уровне средней точности измерений (обычно от 3 до 7 мкм), либо использовать некоторую «хорошую» калибровку для подбора такой оценки σ , которая обеспечивает $\chi_{N-J}^2 \sim 1$. (Этим фактически удовлетворяется критерий 4 из разд. 2.)

Уже отмечалась необходимость проведения нескольких циклов калибровки, что позволяет проверить стабильность калибровочных параметров и в этом случае провести их усреднение. Простейшим критерием стабильности может служить отсутствие дрейфа (систематического сдвига) параметров в последовательных калибровках.

Более точный критерий основан на отношении статистик q_k^2 и s_k^2 [63], вычисляемых по данным измерений координат k -го креста $\{u_{ki}, v_{ki}\}_{k=1}^N \prod_{i=1}^L$. Ограничимся для краткости одной из координат $\{u_{ki}\}$. Тогда

$$q_k^2 = \frac{1}{2(L-1)} \sum_{i=1}^L (u_{k,i+1} - u_{ki})^2;$$

$$s_k^2 = \frac{1}{L-1} \sum_{i=1}^L (u_{ki} - \bar{u}_k)^2,$$

$$\text{где } \bar{u}_k = \frac{1}{L} \sum_{i=1}^L u_{ki}.$$

Определим частное $r_k = q_k^2/s_k^2$. Если в течение измерений присутствует систематический сдвиг, то следует ожидать, что $s_k^2 \gg q_k^2$. Пользуясь таблицами квантилей r_{kp} порядка p распределения случайной величины r_k [63] (обычно $p = 0,05$), при $r_k < r_{kp}$ отмечают нарушение гипотезы стабильности для k -го креста.

Аналогичные статистические вычисления необходимо выполнить для координаты $\{v_{ki}\}$ и для всех крестов. При отсутствии случаев

выхода r_k за пределы критической области для всех крестов $k = 1, N$ можно считать, что результаты калибровочных измерений не зависят от времени.

Причины нарушения стабильности связаны, главным образом, с ремонтом оборудования, сезонной сменой температуры и температурным дрейфом, вызванным разогревом прибора после включения. Последний фактор, сказывающийся иногда в течение 6—10 ч, серьезно ограничивает время функционирования прибора при сменной работе [64, 65].

Исследования, проведенные в этой связи (см., например, [64]), показали, что необходимо выделить калибровочные параметры, определяемые дрейфом измеряемых координат и выбрать интервал времени (обычно 1 ч), через который проводить калибровочные измерения с тем, чтобы в промежутках использовать линейную интерполяцию для корректировки этих существенных параметров.

6. ПРОБЛЕМЫ ПРОГРАММНОЙ РЕАЛИЗАЦИИ

За последние годы было опубликовано немало описаний калибровочных программ для различных автоматизированных измерительных приборов и систем и бесфильмового съема. Отметим, например, [4, 14, 15, 21, 22, 66—69]. За исключением простых программ, предназначенных для полуавтоматической обработки измерительных столов [22] и реализованных на управляющих мини-ЭВМ, подавляющее большинство калибровочных программ работают на больших ЭВМ в off-line режиме. Несмотря на специфику конкретных приложений, попробуем выделить некоторые общие моменты, характерные для таких программ.

Структура данных. Исходными данными являются результаты сканирования калибровочной пластинки (2—4 тыс. точек), идеальные координаты решетки, а в случае двухкоординатных систем типа СИ добавляются ручные измерения во второй системе координат. Массивы исходных данных записываются на магнитную ленту в виде файлов, состоящих из записей стандартной длины и снабженных паспортами, содержащими сведения о приборе, операторе, проводившем измерения, дате и времени измерения и т. д.

Исходные данные обычно требуют перекодировки из формата управляющей мини-ЭВМ в формат большой ЭВМ, на которой ведется обработка. Если есть необходимость экономии памяти, а перекодировка ведется быстро, то в памяти хранят исходный массив без перекодировки, которая выполняется, когда требуется извлечь то или иное измерение по его индексу — пойнтеру в массиве. Во многих программах вводится специальный внутренний формат с плавающей запятой, в который переводятся исходные данные с такой нормировкой, чтобы они были близки к наглядным единицам измерения (миллиметры, радианы и т. д.).

Следует отметить важность контроля исходных данных. Последовательный контроль в программах ввода, перекодировки и т. д. позволяет не тратить время на обработку ложной информации, избежать искажения выходных параметров, оперативно установить дефект измерительной системы. Виды такого контроля: а) геометрический (по монотонности одной из координат, по заданным расстояниям между крестами); б) по-битовый; в) по числу точек, попадающих в заданную область (в одном кресте должно быть не менее 20 точек, на одном плече — не менее 10) и т. д.

Выходные данные — это коэффициенты прямого и обратного преобразования (например, при пятой степени полиномов — это 42 числа), карта поправок (до 600 чисел), значения нормированной величины χ^2_{N-J} и χ^2 для отдельных крестов (еще 300 чисел).

Проверка выходных данных также очень важна. В разд. 4 и 5 отмечалась необходимость вывода на печать различных гистограмм, величин \bar{a} , σ_d , $\max_k d_k$ и контроля за тем, чтобы все эти параметры лежали в определенных допусках. Применяются также проверки комбинаций калибровочных параметров, например масштабов по осям (4), параметра косоугольности (6) и т. д.

Архитектура калибровочной программы. Она обусловлена этапами обработки: получение центров; получение коэффициентов преобразования; вычисление точностных характеристик; контрольная обработка дополнительного эталонного изображения.

Все перечисленные этапы повторяются несколько (10—15) раз, чтобы устранить случайные выбросы и оценить погрешность карты поправок. В калибровочную программу входят также многочисленные подпрограммы проверки и сервисные программы печати, гистограммирования и т. д.

В качестве примера приведем блок-схему, изображенную на рис. 10.

Примеры стандартных калибровочных программ. Ниже приведены сведения о нескольких стандартных программах калибровки, разработанных с участием авторов и употребляемых в Лаборатории вычислительной техники и автоматизации ОИЯИ. Эти программы реализуют алгоритмы, описанные в предыдущих главах для выполнения всех этапов калибровки, и предназначены для приборов, выполняющих параллельно сканирование по всему полю решетки типа тех, что изображены на рис. 1, а, б.

Первый этап выполняется с помощью фортранной программы NIGEN [15], выдающей массив центров крестов и их погрешностей. На втором этапе эти массивы используются для получения коэффициентов прямого и обратного преобразований. В одной из первых стандартных программ калибровки это делалось в программах GENCAL (прямое) и ORCAL (обратное преобразование) [14], использующих системы полиномов $\{\varphi_i^{(1)}\}$ — $\{\varphi_i^{(3)}\}$. В дальнейшем эти программы усовершенствованы и объединены в программу CALORT,

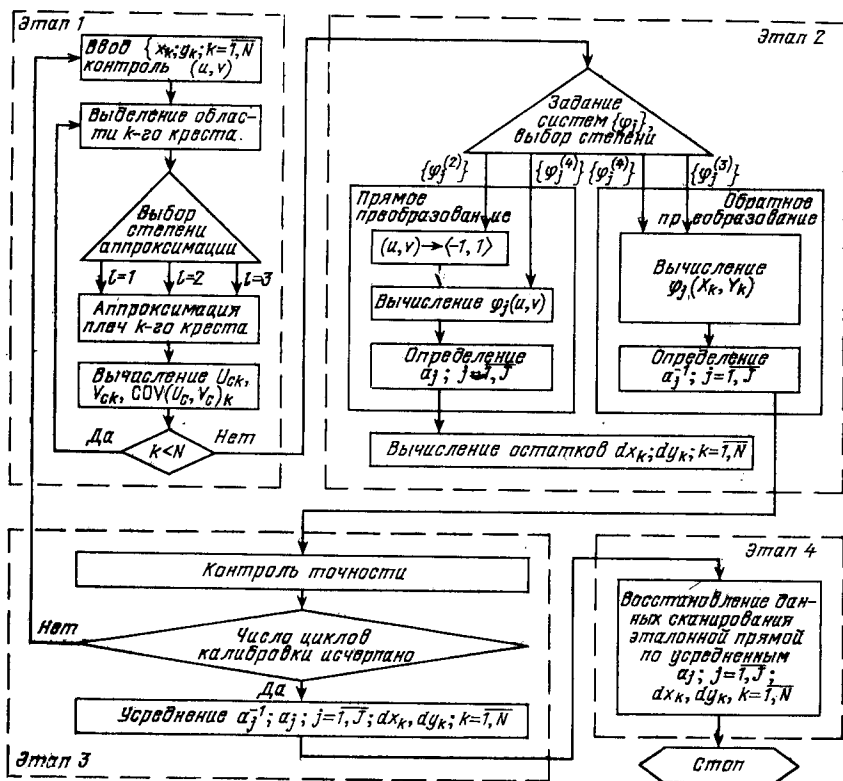


Рис. 10. Принципиальная блок-схема программы калибровки

использующую весь набор систем полиномов $\{\varphi_i^{(1)}\} - \{\varphi_i^{(4)}\}$ [59].

В настоящее время программа NIGEN дополнена программой NORTCR, реализующей алгоритмы нахождения центров крестов на основе аппроксимации одномерными ортонормированными полиномами [54].

Объем оперативной памяти ЭВМ и время счета для программы NIGEN зависят от числа крестов и ключей, задающих выходную печать. Типичный счет с 36 крестами и подробной печатью требует 20 К слов и 30 с счета на ЭВМ CDC-6500.

Аналогично для программ CENCAL и ORCAL требуется по 25 К слов и 20 с времени CDC-6500 для получения коэффициентов прямого и обратного преобразований.

Подробное описание этих программ дано в [14]. Программа CALORT использует ту же вызываемую последовательность, что и GENCAL, но значения параметров NPOL и IOR дополнены для вызова системы полиномов $\{\varphi_i^{(4)}\}$ [59].

Существует два вида программ калибровки автоматов типа спиральный измеритель. Программа SCALP [69] предназначена для семейства приборов Spiral Reader SAAB, имеющего конструктивные отличия от СИ ОИЯИ. Для калибровки последнего разработана калибровочная программа CALIBR, подробное описание которой приведено в [21].

Программу на алгоритмическом языке для символьных операций REDUCE-2 для генерирования двумерных ортогональных полиномов системы $\{\varphi_i^{(2)}\}$ с числом коэффициентов до $J = 50$, а также пример выдачи этой программы можно найти в [41, 45].

В заключение можно привести несколько общих цифр, касающихся средних показателей, относящихся к калибровочным программам.

В общей массе программ обработки основную часть (свыше 50%) составляют программы геометрической реконструкции, кинематического анализа и статистические программы, работающие с лентами суммарных результатов. Программы калибровки и фильтрации примерно равны по объему и вместе составляют оставшиеся 45—50%.

Точность измерения координат по пленке зависит от конструкции автомата и может лежать в пределах 1,6—3 мкм. Такие факторы, как точность пространственных координат, «ложная» кривизна, а также общая скорость обработки, зависят от остальных программ обработки и титульных данных, определяемых экспериментальной установкой.

ЗАКЛЮЧЕНИЕ

Подвергаемые калибровке измерительные системы прошли уже длительную эволюцию, основанную на развитии методов физики высоких энергий, а также измерительной и вычислительной техники. Построение и эксплуатация сложных систем сканирования, в особенности систем с электронно-механической или чисто электронной разверткой, выявили — наряду с их высокой производительностью — существенную нелинейность. Необходимость в учете нелинейности в условиях относительно высокого уровня шумов стимулировала разработку математического аппарата, сочетающего оптимальную обусловленность систем ортонормированных полиномов с робастными методами статистического отсева шумовых точек. Этот математический аппарат находит применение на обоих этапах калибровки: определения координат реперных крестов и построения преобразований калибровки [70]. Составленные при помощи такого аппарата алгоритмы и программы являются свидетельством его работоспособности и практической применимости. Представляется перспективным внедрение быстрых алгоритмов счета в уже существующие программы.

В ОИЯИ разработаны стандартные пакеты калибровочных программ GENCAL, ORCAL и CALORT. Они были использованы в калибровке разнообразных систем и, по-видимому, будут применяться и в будущем. В этой связи необходимо подчеркнуть исключительную

важность учета конкретных особенностей калибруемой системы на основе принципа априорной информации А. Н. Тихонова [71], а также необходимость тщательной проверки данных на всех этапах калибровки. Только при соблюдении указанных условий можно ожидать разумные результаты, способные служить основой как для правильной интерпретации измеряемых величин, так и для дальнейшего совершенствования приборов и систем.

СПИСОК ЛИТЕРАТУРЫ

1. Автоматическая обработка данных с пузырьковых и искровых камер; Пер. с англ./Под ред. Б. Г. Розова. М.: Атомиздат, 1974.
2. Комлев В. П. и др. Препринт ИТЭФ № 69. М., 1974.
3. Котов В. М. и др. ОИЯИ 10-7939. Дубна, 1974.
4. Proc. of European Spiral Reader Symposium, CERN 72-46. Stockholm, 1972.
5. Стрэнд Р. Распознавание образов при помощи ЦВМ: Пер. с англ. М.: Мир, 1974.
6. Proc. of Intern. Conf. on Data Handling Systems in High-Energy Physics. CERN 20-21. Cambridge, 1970.
7. Proc. of Conf. on Computer Assisted Scanning, Padova, 1976.
8. Материалы II Всесоюзного семинара по обработке физической информации. Ереван: изд. ЕрФИ, 1978.
9. Garvey J. e. a. CERN-NP/OM-SN. 140, Geneva, 1973.
10. Алексеева Н. П. и др. ОИЯИ Б1-43-80-133. Дубна, 1980.
11. Байла И. и др. ОИЯИ P10-82-653. Дубна, 1982.
12. Крушинский Д. и др. Материалы V Международного совещания по проблемам математического моделирования и программирования. ОИЯИ P10, 11-84-818. Дубна, 1984. С. 337—339.
13. Klein F., Ströbele H.//Proc. of Intern. Conf. on Data Handling Systems in High-Energy Physics. Cambridge, 1970. CERN 70-21. P. 177—183.
14. Байла И., Ососков Г. А. ОИЯИ P10-11834. Дубна, 1978.
15. Алексеева Н. П. и др. ОИЯИ 10-81-483. Дубна, 1981.
16. Gadjokov V. JINR-E11-80-782. Dubna, 1980.
17. Горчаков О. Е. ОИЯИ P11-84-188. Дубна, 1984.
18. Anderson M. e. a. SLAC-PUB-599. Stanford, 1960.
19. Тихонов А. Н. Математические модели и научно-технический прогресс (автоматизация обработки наблюдений)//Наука и человечество — 1979. М.: Знание, 1979. С. 283.
20. Davey P. G.//Proc. on Comp. Assited Scanning. Padova, 1976. P. 93—113.
21. Мороз В. И., Ососков Г. А. ОИЯИ P10-84-791. Дубна, 1984.
22. Калмыкова Л. П. и др. ОИЯИ 10-8808. Дубна, 1975.
23. Айвазян С. А. и др. Прикладная статистика. М.: Финансы и статистика, 1985.
24. Ососков Г. А. ОИЯИ P11-4463. Дубна, 1969.
25. Арешидзе Г. Г. и др.//Труды ТГУ. Т. 208. Вып. 7. Тбилиси: Изд. ТГУ. 1979. С. 57—61.
26. Ососков Г. А. ОИЯИ P10-83-187. Дубна, 1983.
27. Смоляк С. А., Титаренко Б. П. Устойчивые методы оценивающие. М.: Статистика, 1980.
28. Хьюбер П. Робастность в статистике: Пер. с англ. М.: Мир, 1984.
29. Куняев С. В. и др. ОИЯИ P10-84-553. Дубна, 1984.
30. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия: Пер. с англ. М.: Финансы и статистика, 1977.
31. Воеводин В. В., Кузнецов Ю. А. Матрицы и вычисления. М.: Наука, 1984. § 5.
32. Крылов В. И. Приближенное вычисление интегралов. М.: Наука, 1967.

33. Мысовских И. П. Интерполяционные кубатурные формулы. М.: Наука, 1981.
34. Кнут Д. Искусство программирования на ЭВМ. Т. 2: Пер. с англ. М.: Мир, 1977.
35. Forsythe G. E. // J. Soc. Ind. Appl. Math. 1957. Vol. 5. P. 74—88.
36. Weisfeld M. Numerische Mathematik. 1959. Bd 1. S. 38—40.
37. Гаджиков В., Богданова Н. ОИЯИ-Р11-42860. Дубна, 1979.
38. Gadjokov V., Jordanova J. // Comp. Phys. Com. 1984. Vol. 31. P. 53.
39. Rosato E. e. a. // Comp. Phys. Com. 1981. Vol. 22. P. 311—315.
40. Корн Г., Корн Т. Справочник по математике: Пер. с англ. М.: Наука, 1970.
41. Байла И., Ососков Г. А. Сборник научных работ ОИЯИ-ЦИФИ. Т. 3. Будапешт, 1982. КФКИ 1979-82. С. 47.
42. Ascher M., Forsythe G. E. // J. Assoc. Comp. Mach. 1958. Vol. 5. P. 9.
43. Beckmann P. Orthogonal Polynomials for Engineers and Physicists. The Golem Press. Boulder. Co. (USA), 1973.
44. Лаврентьев М. А., Шабат Б. В. Методы теории функций комплексного переменного. М.: Наука, 1973.
45. Байла И., Ососков Г. А., Хэрн А. К. ОИЯИ-Р10-11944. Дубна, 1978.
46. Bohm G. e. a. // Proc. of the 1st Intern. Conf. on Streamer Chamber Technology, ANL, 1972. P. 117.
47. Bogdanova N., Gadjokov V. // Comp. Phys. Comm. 1981. Vol. 24. P. 225.
48. Гаджиков В., Богданова Н. ОИЯИ-Р11-80-781. Дубна, 1980.
49. Худсон Д. Статистика для физиков: Пер. с англ. М.: Мир, 1967.
50. Friebel W. Report PHE-80-9, Zeuthen, 1980.
51. Vojkov G., Gadjokov V., Ilieva K. Report IAEA-INDC (BUL)-6/GV, Vienna, 1982; см. также Ядрена Энергия. 1983. Т. 18. С. 40.
52. Vojkov G., Gadjokov V., Minchev S. Report IAEA-INDC(BUL)-007/GV, Vienna, 1983.
53. Косарев Е. Л. Инф. бюл. Алгоритмы и программы. Вып. 1. Программа № P000257. М., 1973.
54. Богданова Н. Б. и др. ОИЯИ 10-83-93. Дубна, 1983.
55. Ососков Г. А. ОИЯИ 5-9787. Дубна, 1976.
56. Богданова Н. Б. и др. // Болг. физ. журн. 1984. Т. 11. Вып. 4. С. 387; Богданова Н. // Математическое моделирование в технике. Варна, 1982. С. 387—398.
57. Дунин-Барковский И. В., Смирнов Н. В. Теория вероятностей и математическая статистика. М.: ГИИТЛ, 1955.
58. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1974.
59. Богданова Н. Б. и др. Материалы V Международного совещания по проблемам математического моделирования и программирования. ОИЯИ P10, 11-84-848. Дубна, 1984. С. 323—326.
60. Bogdanova N. V. JINR E11-84-512. Dubna, 1984
61. Абдурахимов А. У. и др. ОИЯИ 10-80-831. Дубна, 1980.
62. Статистические методы в экспериментальной физике: Пер. с англ. / Под ред. А. А. Тяпкева. Прил. 3. М.: Атомиздат, 1976. С. 319.
63. Ливник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений. М.: Физматгиз, 1962.
64. Войтенко Ю. Г. и др. ОИЯИ P10-10610. Дубна, 1977.
65. Байла И. и др. ОИЯИ P10-12990. Дубна, 1980.
66. Аникеев В. Б. и др. ИФЭЭ, ОМТВ 75-91. Серпухов, 1975.
67. Астахов А. Я. и др. ОИЯИ 10-10837. Дубна, 1977.
68. Карлов А. А., Сельченко В. А. ОИЯИ 10-11155. Дубна, 1977.
69. Закамский Л. Л. и др. ИФЭЭ, ОЭА 77-132. Серпухов, 1977.
70. Bogdanova N. // Journ. of Comp. and Appl. Math., Belgium, 1986, Vol. 14. P. 345—351.
71. Тихонов А. Н. и др. Регуляризирующие алгоритмы и априорная информация. М.: Наука, 1983.