

Обзор методов обратной свёртки

Ю.В. Богомолов, В.В. Алексеев, О.А. Леванова, А.Г. Майоров, В.В. Малахов

Сегодня аппаратура для проведения экспериментальных измерений физических величин в самых разных научных направлениях обладает очень высокими точностью и чувствительностью. Однако, как оказывается, исключить полностью влияние инструментальных эффектов на результат невозможно. Изменяемые значения физической величины неизбежно будут отличаться от истинных и в некоторых случаях даже значительно. Таким образом, возникает задача восстановления истинных распределений из измеренных с учётом особенностей проведения эксперимента и характеристик научной аппаратуры. Применяются разные подходы, основанные на математической модели прибора и постановке задачи обратной свёртки. Представлено описание этой задачи, ключевых идей и методов её решения, а также особенностей и деталей реализации на примере экспериментов в области физики элементарных частиц и космофизики.

Ключевые слова: обратная свёртка, деконволюция, анфолдинг, методы восстановления спектра, статистические оценки, байесовские методы, регуляризация, безбиновые методы, машинное обучение

PACS numbers: 02.50. – r, 02.60. – x, 96.50.S – , 96.50.sb

DOI: <https://doi.org/10.3367/UfNr.2022.05.039189>

Содержание

1. Введение (669).
 2. Задача обратной свёртки (670).
 - 2.1. Общие положения. 2.2. Используемые обозначения и постановка задачи в случае дискретизации. 2.3. Прямой метод решения задачи обратной свёртки.
 3. Методы решения задачи обратной свёртки (672).
 - 3.1. Байесовские методы. 3.2. Методы, основанные на регуляризации. 3.3. Метод поправочных коэффициентов. 3.4. Прочие методы восстановления спектра. 3.5. Программные реализации.
 4. Использование методов обратной свёртки в экспериментах по физике частиц (679).
 5. Перспективы развития методов обратной свёртки. Заключение (681).
- Список литературы (682).

1. Введение

Уровень современных экспериментов в разных научных направлениях становится всё выше, научная аппаратура чувствительнее, а получаемые результаты точнее. Благодаря этому совершаются новые открытия, изучаются недоступные ранее эффекты и явления, уточняется наше представление о законах природы. Вместе с тем неиз-

бежно возрастают требования к качеству проведения обработки и анализа получаемых данных, чтобы устранить влияние на результат особенностей работы прибора. Не являются исключением и области физики атомного ядра, элементарных частиц, космофизики и других смежных дисциплин.

В работе рассматривается одна из серьёзных задач, возникающих при обработке научных данных современных экспериментов в указанных направлениях, — восстановление истинных значений характеристик частиц и их распределений из измеренных величин.

Неточности измерительных приборов и шумы аппаратуры приводят к тому, что характеристики частиц, такие как скорость и магнитная жёсткость, энергия и импульс, направление движения и др., подвергаются различным искажениям. Как следствие, полученные в эксперименте распределения указанных характеристик, например энергетические спектры, угловые и массовые распределения, будут отличаться от истинных, что может приводить к неверной интерпретации результатов измерений и ложным физическим выводам.

Такая проблема может быть решена с помощью так называемых методов обратной свёртки, которые позволяют восстановить истинное распределение рассматриваемой характеристики по измеренным данным (точнее, его оценку, максимально приближённую к истинному, но, вообще говоря, неизвестному закону). Эти методы используют сведения о предполагаемом характере искажений во время измерения, которые могут быть получены с помощью моделирования отклика прибора методами Монте-Карло или в результате калибровки. На основе подобных данных, для которых известны истинные и измеренные значения физической величины, строится матрица миграций [1, 2]. Свёртка матрицы миграций с истинным распределением является измеренным распределением этой величины. При обработке

Ю.В. Богомолов^(1,2,a), В.В. Алексеев^(1,2), О.А. Леванова^(1,2,b),
А.Г. Майоров⁽²⁾, В.В. Малахов⁽²⁾

⁽¹⁾ Ярославский государственный университет им. П.Г. Демидова,
ул. Советская 14, 150003 Ярославль, Российская Федерация

⁽²⁾ Национальный исследовательский ядерный университет МИФИ,
Каширское шоссе 31, 115409 Москва, Российская Федерация
E-mail: ^(a) mathematics@inbox.ru, ^(b) olaydy@gmail.com

Статья поступила 23 января 2022 г.,
после доработки 7 апреля 2022 г.

экспериментальных данных интерес представляет решение обратной задачи, однако свёртка обратной матрицы миграций с измеренным распределением является неустойчивой оценкой истинного распределения [3–6]. Поэтому для получения устойчивого решения используются алгоритмы обратной свёртки, основанные на различных математических подходах.

В физике частиц применяются различные статистические методы [5, 7]. Обсуждение подходов к решению задачи обратной свёртки, особенностей конкретных алгоритмов и их приложений к обработке результатов экспериментов в физике высоких энергий активно ведётся с начала 1980-х годов. В 1984 году общий обзор алгоритмов для решения этой задачи проведён Блобелем в рамках проводимой ЦЕРНом летней школы компьютерных вычислений и представлен в работе [1]. В 1998 году подробное описание некоторых методов обратной свёртки приведено в соответствующей главе монографии Коуэна [8], посвящённой статистическому анализу данных, а также в его обзорной статье [5]. Описание актуальных (на тот момент) методов обратной свёртки приведено в более поздней обзорной статье Блобеля [9] (2011 г.), а также в написанной им главе практического руководства по применению статистических методов в области физики высоких энергий [10].

Отмеченные обзорные публикации дают хорошее представление о постановке проблемы восстановления спектра физических величин, возможных ошибках при использовании "наивного" подхода к решению задачи, а также предоставляют подробное описание (от идеи до алгоритма) некоторых методов решения задачи обратной свёртки. К недостаткам этих работ следует отнести значительный перекос в сторону описания регуляризационных методов восстановления дискретного спектра и соответственно недостаточное освещение других методов обратной свёртки. Также можно отметить отсутствие в указанных публикациях описаний практического использования соответствующих алгоритмов. Более широкий обзор методов обратной свёртки предложен в работах [11–15]. Также важное прикладное значение имеют работы [16, 17], в которых помимо обзора методов приводится сравнительный анализ алгоритмов, реализованных в программных пакетах обратной свёртки, активно применяемых при обработке результатов физических экспериментов.

Описание алгоритмов обратной свёртки встречается в обзорной части или при перечислении методов обработки данных в диссертационных работах. В некоторых из них авторы ограничиваются упоминанием конкретного выбранного метода [18, 19], но могут и включать сравнительный анализ алгоритмов обратной свёртки [20], в том числе достаточно подробный обзор [21].

В настоящее время методы обратной свёртки продолжают развиваться: модифицируются существующие и разрабатываются новые алгоритмы. В частности, активно применяются методы машинного обучения, хорошо зарекомендовавшие себя в различных прикладных областях. Такие подходы оказываются применимыми в том числе для построения оценок непрерывного спектра с помощью так называемых безбиновых методов обратной свёртки, возможности использования которых в прикладных задачах в последнее десятилетие активно обсуждаются [22–25].

В настоящей работе приводится широкий обзор методов обратной свёртки для восстановления дискретных и непрерывных спектров: подробно рассматриваются самые основные и часто используемые методы обратной свёртки, отмечаются их преимущества и недостатки, а также обсуждаются некоторые общие вопросы, важные при обработке результатов экспериментальных измерений.

2. Задача обратной свёртки

В этом разделе обсуждается общий подход к восстановлению оценки истинного распределения физической величины, вводятся основные термины и обозначения, рассматриваются базовая постановка задачи и прямой метод её решения.

2.1. Общие положения

Несовпадение измеренных и истинных значений физической величины может быть как малым, так и значительным, в общем случае приводя к отличию экспериментальных распределений от истинных. Выделяют несколько причин, по которым величины измеряются неточно и отличаются от истинных значений. Чаще всего это происходит из-за конечного разрешения измеряющей аппаратуры или шумов, а также под влиянием различных физических процессов, связанных с прохождением частиц через вещество прибора, например рассеяния или потерь энергии.

Таким образом, возникает необходимость восстановления истинного распределения измеряемой величины по измеренному, которая может быть сформулирована в виде задачи в рамках подхода, известного как обратная свёртка. Также в литературе его можно встретить под названиями деконволюция или *анфолдинг* (от англ. *unfolding* — разворачивание).

В некоторых случаях формулировка задачи может быть упрощена, например, если известен общий вид искомого распределения и можно оценить один или несколько входящих в него параметров. Однако в общем случае закон распределения неизвестен, кроме, может быть, его отдельных свойств, таких как непрерывность или гладкость. Это требует применения непараметрических методов математической статистики. В частности, для оценки плотности распределения используются: гистограммные методы, предполагающие разбиение множества значений величины на промежутки; методы аппроксимации плотности по некоторой системе базисных функций; методы локального оценивания плотности по выборке и др.

Подлежащий восстановлению закон распределения значений физической величины может быть непрерывным или дискретным. Обычно в решаемых научных задачах встречается именно первый вариант, и тогда наиболее широко применяется гистограммный метод с дискретизацией непрерывного спектра физической величины. Для оценки истинного закона распределения измеряемой величины в этом случае используется следующий подход:

- множество значений измеряемой величины разбивается на интервалы (бины);
- для каждого интервала подсчитывается количество попавших в него событий;

- строится приборная гистограмма, являющаяся дискретным приближением непрерывного распределения измеряемой величины;

- определяется функция отклика прибора, учитывающая отличие измеренных значений величины от истинных (матрица миграций);

- строится статистическая оценка истинного распределения.

Необходимо отметить, что результатом восстановления истинного распределения в данном случае является не непрерывная функция, а дискретная статистическая оценка, в любом случае отличающаяся от дискретизации истинного неизвестного распределения физической величины. Однако использование подхода обратной свёртки позволяет получить более точную статистическую оценку, приближенную к истинному закону.

Прочие подходы к решению задачи восстановления законов распределения физической величины, не предполагающие выделения такой системы интервалов значений, можно условно отнести к безбиновым (unbinned, binning-free) методам. В этом случае алгоритм восстановления спектра основан на конкретной идее оценивания неизвестной плотности распределения. Выделяют следующие основные подходы:

- ядерное сглаживание [26] — оценка неизвестной непрерывной плотности распределения строится как среднее арифметическое значений функции специального вида (ядерной функции) в точках выборки;

- проекционный метод [27] — непрерывная плотность распределения оценивается как линейная комбинация некоторой системы (базиса) функций;

- сплайн-подход [28] — плотность распределения оценивается как кусочно заданная функция, компоненты которой на отдельных промежутках оцениваются по выборке;

- локальный метод — значение плотности распределения оценивается только в окрестности конкретной выбранной точки (используются метод k -ближайших соседей [29], метод потенциальных функций [30] и другие подходы).

Отметим также, что существуют методы, которые применяются в обоих вариантах — как для построения дискретных оценок (гистограмм) плотности на выделенной системе интервалов, так и для сглаженной оценки плотности распределения, описанные, например, в статье [24] (см. также приведённые там ссылки).

2.2. Используемые обозначения и постановка задачи в случае дискретизации

Рассмотрим некоторую физическую величину со значениями, распределёнными по неизвестному закону. Как отмечалось в разделе 2.1, наиболее распространённым вариантом оценивания закона распределения является построение оценок вероятностей попадания значений случайной величины в некоторый конечный набор интервалов. В этом случае диапазон её возможных истинных значений можно разбить на множество бинов $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_n)$, каждому из которых ставится в соответствие вероятность нахождения в нём истинного значения величины. Процедуру такого разбиения в дальнейшем будем называть биннингом, а полученный набор вероятностей $p = (p_1, p_2, \dots, p_n)$ — дискретизацией непрерывного распределения или истинным спектром, оценку которого необходимо получить.

Разбиение в общем случае может быть произвольным, неравномерным, однако от выбора биннинга существенно зависит качество восстановления оценки истинного распределения (этот вопрос детальнее обсуждается ниже).

Пусть в ходе эксперимента зарегистрировано N частиц, для каждой из которых измерено значение исследуемой величины; такой единичный факт называют событием. Математическое ожидание истинного количества событий в интервалах разбиения, которое обозначим как $\tau = (\tau_1, \tau_2, \dots, \tau_n)$, вычисляется по формуле $\tau = Np$.

Диапазон измеренных значений величины также разбивается на бины $\Delta' = (\Delta'_1, \Delta'_2, \dots, \Delta'_k)$. Отметим, что допустимо использование разного биннинга для истинных и измеренных значений, в том числе таких, что $n \neq k$. Количество зарегистрированных частиц в соответствующих бинах обозначим $m = (m_1, m_2, \dots, m_k)$ и будем называть измеренным спектром. В некоторых случаях под ним также понимают соответствующие частоты $m/N = (m_1/N, m_2/N, \dots, m_k/N)$.

На рисунке 1 иллюстрируются истинный и измеренный спектры некоторой произвольной величины. Разница между ними, обусловленная указанными выше причинами, может быть значительной для отдельных интервалов или их последовательности.

Описать различие спектров можно с помощью матрицы миграций R , где элемент R_{ij} — вероятность того, что истинное значение величины из j -го бина будет зарегистрировано в i -м бине. Тогда математические ожидания $v = (v_1, v_2, \dots, v_k)$ количества событий со значениями характеристик, лежащих в рассматриваемых интервалах, вычисляются следующим образом: $v = R\tau$.

На рисунке 2 приведён иллюстративный пример матрицы миграций при разном разбиении на интервалы по осям истинного и измеренного значений физической величины, цветовая шкала показывает значения R_{ij} матрицы миграции.

Таким образом, основной задачей является разработка методов состоятельной статистической оценки неизвестных p или пропорциональных им математических ожиданий τ по измеренному спектру m при известном характере искажений, описываемом с помощью матрицы миграций R .

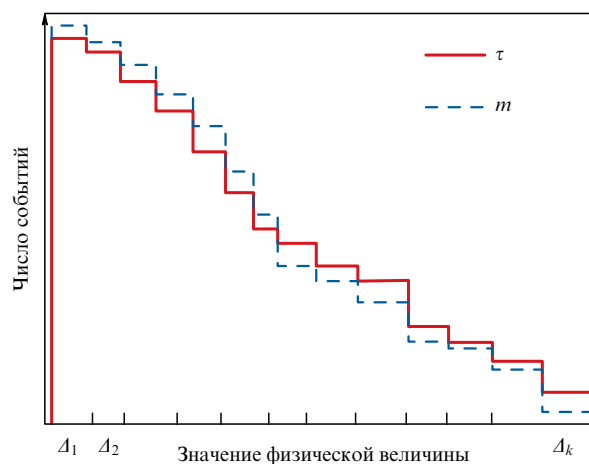


Рис. 1. Иллюстрация истинного (сплошная линия) и измеренного (штриховая линия) распределений некоторой физической величины.

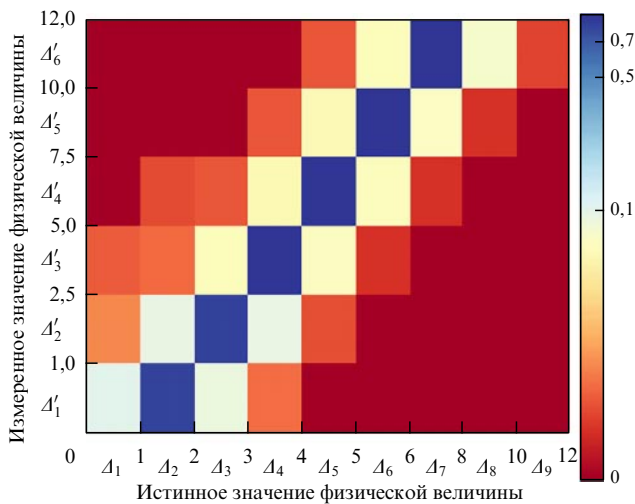


Рис. 2. Иллюстративное представление матрицы миграций.

2.3. Прямой метод решения задачи обратной свёртки

Для решения задачи естественным является допущение, что события не могут произойти одновременно (условие ординарности). Вероятность регистрации фиксированного количества событий m_i в интервале Δ_i в течение времени наблюдения не зависит от наступления событий в предыдущие моменты (отсутствие последствия) и от начала отсчёта времени наблюдения (условие стационарности). При указанных допущениях выполняются все необходимые условия для того, чтобы считать последовательность событий простейшим (пуассоновским) потоком [2].

В последнем случае количество событий, попавших в интервал Δ_i , является случайной величиной, распределённой по пуассоновскому закону с математическим ожиданием v_i . Тогда вероятность наступления m_i таких событий

$$P(m_i, v_i) = \exp(-v_i) \frac{v_i^{m_i}}{m_i!}.$$

Таким образом, рассматривается набор случайных величин с пуассоновскими распределениями, зависящими от неизвестных математических ожиданий v , которые в свою очередь зависят от неизвестного τ ($v = R\tau$). Для построения состоятельной статистической оценки τ по выборке m может использоваться метод максимального правдоподобия. Для этого составляется и максимизируется (по τ) функция правдоподобия:

$$L(\tau) = \prod_i P(m_i, v_i) = \prod_i \exp(-v_i) \frac{v_i^{m_i}}{m_i!} \rightarrow \max_{\tau}.$$

Максимум построенная функция правдоподобия $L(\tau)$ достигает при $v_i = m_i$, а оценкой τ является $\hat{\tau} = R^{-1}m$ [3, 5]. Это соответствует "наивному подходу" к решению задачи обратной свёртки: считается, что измеренное и истинное распределения связаны линейным матричным уравнением, которое решается применением обратной матрицы миграций к измеренному спектру.

Построенная точечная оценка $\hat{\tau}$ является состоятельной и несмещённой [5, 31–33]. Однако на практике использование прямого метода возможно лишь при наличии выборки большого объёма, а также очень высокого разрешения измеряющей аппаратуры, что соответствует матрице R , близкой к диагональной [3, 5]. В про-

тивном случае оценка является чувствительной к малым возмущениям измеренного распределения (т.е. прямая оценка неустойчива), что приводит к неприемлемо большой ошибке (неопределённости) при оценке τ [3–6]. В то же время указать точные границы применимости прямого метода, где он остаётся устойчивым, нельзя, поскольку в общем случае это зависит от большого числа факторов, уникальных для каждого эксперимента, и в каждом случае требуется своё исследование.

Помимо отмеченной неустойчивости, описанный подход в исходном виде не позволяет учесть особенности истинного распределения, например его гладкость или энтропию [8, 34].

Совокупность указанных недостатков не позволяет широко использовать прямой метод для восстановления оценки истинного распределения, поскольку он не гарантирует получения корректного устойчивого результата. Возможным вариантом решения этой проблемы является искусственное введение смещения статистической оценки, которое может быть выполнено двумя наиболее распространёнными способами [35]:

- 1) добавление "штрафного" слагаемого в функцию правдоподобия (на чём основаны использующие регуляризацию Тихонова методы, обсуждающиеся далее) [36, 37];
- 2) ранняя остановка итеративных алгоритмов построения статистических оценок [38].

В разделе 3 подробно обсуждаются специальные методы решения данной задачи, основанные на разных математических алгоритмах обратной свёртки и отмеченных идеях построения смещённых оценок.

3. Методы решения задачи обратной свёртки

Для уменьшения статистических ошибок и сглаживания оценки истинного распределения используют как некоторые модификации прямого метода, так и принципиально другие подходы к обратной свёртке, среди которых можно отметить следующие:

- 1) байесовские подходы к решению задачи обратной свёртки [39] (методы д'Агостини [38, 40], модификации Куусела, Панаретоса [41, 42]), алгоритм FBU (Fully Bayesian Unfolding) [43];
- 2) регуляризация Тихонова и её модификации (метод сингулярного разложения (Singular Value Decomposition, SVD): Картвелишвили, Хёккер [36, 44], регуляризация, основанная на энтропии [5], реализованный в программном пакете ROOT метод TUnfold [12, 37]);
- 3) введение поправочных коэффициентов и систематических сдвигов (модельная зависимость) [5];
- 4) методы машинного обучения.

3.1. Байесовские методы

Одним из распространённых подходов к решению задачи обратной свёртки является идея использования формулы Байеса для оценки вероятностей попадания частиц в интервалы по частотам наблюдаемых событий и имеющейся матрице миграций. Опишем особенности применения данного подхода к задаче восстановления спектра физической величины.

Будем обозначать события, связанные с реальным попаданием физической величины в соответствующие интервалы $\Delta_1, \Delta_2, \dots, \Delta_n$ как $C_1, C_2, \dots, C_n$. События, связанные с детектированием события в тех же интервалах,

обозначим через $E_1, E_2, \dots, E_n$. Отметим, что вероятности $P(C_1), P(C_2), \dots, P(C_n)$ неизвестны — это компоненты интересующего нас истинного спектра ($p_i = P(C_i)$), а вероятности, с которыми частицы детектируются в интервалах, могут быть оценены по результатам эксперимента как $\hat{P}(E_j) = m_j/N$. Также, используя данные калибровок или численного моделирования, можно оценить условные вероятности детектирования характеристики в каждом j -м интервале при условии реального попадания значения в i -й интервал. Эти условные вероятности ранее были проинтерпретированы как компоненты матрицы миграции $R_{ij} = P(E_j|C_i)$.

Таким образом, процедура обратной свёртки переформулируется в терминах событий и их вероятностей как задача восстановления неизвестных вероятностей $P(C_i)$ или пропорциональных им математических ожиданий истинного количества событий в соответствующих интервалах по известным условным вероятностям $P(E_j|C_i)$ и вероятностям $P(E_j)$ (экспериментально оцениваются как $\hat{P}(E_j)$). Для решения этой задачи необходимо оценить условную вероятность реального нахождения физической величины в j -м интервале при условии того, что она детектирована в i -м интервале. Такая оценка проводится по формуле Байеса:

$$M_{ji} = P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{P(E_j)} = \frac{P(E_j|C_i)P(C_i)}{\sum_{k=1}^n P(C_k)P(E_j|C_k)}.$$

Тогда математическое ожидание количества событий, детектированных в j -м интервале (всего таких событий m_j), при условии, что истинное значение величины лежит в i -м интервале, оценивается как $m_j P(C_i|E_j) = M_{ji}m_j$. Соответственно математическое ожидание истинного количества событий в i -м интервале получается суммированием по всем интервалам, в которых детектированы события: $\hat{\tau}_i = \sum_{j=1}^n M_{ji}m_j$. Построенную матрицу $M = (M_{ji})$ будем называть матрицей обратной свёртки; в общем виде она не является обратной к матрице миграций R .

На практике необходимо учитывать, что физическая величина из j -го интервала может быть детектирована вне рассматриваемого диапазона. Вероятность регистрации хотя бы в одном из рассматриваемых интервалов называется эффективностью и определяется как $\varepsilon_j = \sum_{k=1}^n P(E_j|C_k)$. В случае $\varepsilon_j \neq 1$ ожидаемое истинное количество событий в j -м интервале вычисляется следующим образом: $\hat{\tau}_i = (1/\varepsilon_i) \sum_{j=1}^n M_{ji}m_j$. В дальнейшем будем считать, что поправка на эффективность включена в матрицу M , а саму матрицу при этом также будем называть матрицей обратной свёртки.

Таким образом, задача оценки истинного спектра (обратной свёртки) сводится к построению матрицы M . Однако в формуле Байеса для оценки элементов матрицы M_{ji} присутствуют вероятности $P(C_i)$, которые изначально неизвестны. Иначе говоря, при таком подходе для оценки истинного спектра уже должно быть известно распределение вероятностей попадания физической величины по соответствующим интервалам (т.е. сам истинный спектр), что делает задачу обратной свёртки некорректной. Оценить вероятности $P(C_i)$ (в отличие от $P(E_j)$) непосредственно по экспериментальным данным нельзя. Поэтому для восстановления истинного спектра используется описанный в работе д'Агостини [38] метод, основанный на подходе, известном в других областях как

алгоритм Ричардсона – Люси [45–48]. Метод основан на следующей идее: выбирается какое-либо начальное приближение истинного спектра, которое в дальнейшем последовательно уточняется. При этом для построения необходимой матрицы M в качестве оценки вероятностей $P(C_i)$ в формуле Байеса используется текущая оценка истинного спектра. В предположении, что такая процедура последовательного уточнения оценки сходится, предельное распределение рассматривается как решение задачи обратной свёртки.

В пакетах ROOT [49] и RooUnfold [50] данная идея реализована в виде итеративного алгоритма:

1. В качестве начального приближения неизвестных вероятностей $P(C_i)$ взять произвольное вероятностное распределение $P_0(C_i)$. Причём желательно выбирать распределение, близкое к истинному (как правило, такое априорное распределение выбирается исходя из известных физических представлений). При отсутствии такой возможности в первом приближении можно взять дискретное равномерное распределение $P_0(C_j) = 1/k$.

2. С помощью формулы Байеса построить оценку истинного распределения:

$$M_{ji} = \frac{R_{ij}P_0(C_i)}{\sum_{k=1}^n \left(P(E_j|C_k) \sum_{l=1}^n R_{kl}P_0(C_l) \right)},$$

$$\hat{\tau}_i = \sum_{j=1}^n M_{ji}m_j, \quad \hat{P}(C_i) = \frac{\hat{\tau}_i}{\sum_{k=1}^n \hat{\tau}_k}.$$

3. Оценить различия между распределением, найденным на шаге 2, и инициализирующим распределением истинной случайной величины (по критерию χ^2 или другому критерию согласия).

4. В качестве нового инициализирующего распределения $P_0(C_i)$ выбрать построенную оценку $\hat{P}(C_i)$, повторять шаги 2–4 до тех пор, пока отклонения между очередной оценкой и инициализирующим распределением не станут достаточно малыми (ниже критического значения).

Для данной реализации метода наиболее существенным недостатком является открытый вопрос о сходимости итерационного процесса. Также отметим, что сходимость итерационного процесса не означает построения достаточно точной оценки: при большом количестве итераций могут накапливаться статистические ошибки, что приводит к большим отклонениям восстановленного спектра от истинного [31]. Напротив, ранняя остановка итерационного процесса при малых изменениях распределений на соседних итерациях, как было отмечено, рассматривается как возможный способ построения смещённых оценок, но с более низкой дисперсией [31, 32, 38]. Поэтому количество итераций, требуемое для приемлемого решения задачи восстановления спектра, приходится находить в ходе серии численных экспериментов с модельными распределениями.

Например, реализованная в пакете RooUnfold версия алгоритма использует следующую основную идею остановки итерационного процесса — в конце каждой итерации рассматриваются два распределения интересующей случайной величины: текущее распределение и построенное на его основе (указанным выше способом) новое распределение. Анализируется гипотеза об однородности упомянутых законов распределения: если гипотеза верна, то шаг итерационного процесса не оказал значи-

мого воздействия на закон распределения случайной величины, поэтому дальнейшие итерации нецелесообразны. Соответственно отклонённую гипотезу об однородности можно интерпретировать как заметное изменение закона распределения (спектра) за текущий шаг и расходимость итерационного процесса. В работах [38, 40] предлагается использовать для проверки этой гипотезы χ^2 -критерий Пирсона (и соответствующую метрику различия между текущим и новым распределением). Количество итераций, необходимое для достижения указанного критерия остановки, заранее неизвестно. Отметим, что реализация из пакета RooUnfold даёт возможность явно указывать количество итераций, позволяя тем самым более гибко управлять остановкой итерационного процесса. В частности, для иллюстрации работы алгоритма авторы пакета используют четыре итерации, схожие результаты получены при обработке данных в эксперименте ALICE (A Large Ion Collider Experiment) [51] на Большом адронном коллайдере (Large Hadron Collider, LHC).

Описанный алгоритм основывается на идее последовательной корректировки вероятностей попадания в каждый рассматриваемый интервал (бин) в отдельности с помощью формулы Байеса. В дальнейшем автором [40] был предложен модифицированный метод, ключевой идеей которого является восстановление распределения количества событий в бинах при предположении, что распределение мультиномиальное. При этом параметры распределения итеративно корректируются также на основе байесовского подхода. Для критерия остановки итерационного процесса используется та же процедура, что в исходном методе. Следовательно, важной задачей является анализ условий, при которых данный критерий остановки процесса приводит к корректному восстановлению распределения, а также условий сходимости итераций.

3.2. Методы, основанные на регуляризации

Опишем подход к решению задачи обратной свёртки, основанный на регуляризации.

Напомним, что прямое решение задачи обратной свёртки чувствительно к малым возмущениям (неустойчиво). Регуляризация заключается во введении дополнительных ограничений к функционалу исходной оптимизационной задачи [52], что не позволяет при малых изменениях в данных строить существенно различающиеся решения.

Таким образом, предлагается максимизировать не исходную функцию правдоподобия (или её логарифм $\ln L(\tau)$), а функцию, включающую в себя дополнительное слагаемое или множитель. Присоединённое слагаемое может являться характеристикой некоторых особенностей исходного распределения (например, гладкости). Фактически задача состоит в нахождении такого распределения, для которого оптимальное значение функции правдоподобия не достигается, но полученная оценка устойчива к возмущениям измеренных распределений.

Итак, задача обратной свёртки может быть записана в виде оптимизационной задачи:

$$\Phi(\tau) = \ln L(\tau) + \alpha S(\tau) \rightarrow \max_{\tau},$$

где α — числовой показатель (параметр регуляризации), описывающий вклад функции $S(\tau)$ (например, отражаю-

щей предполагаемые особенности восстанавливаемого распределения).

Регуляризационную функцию $S(\tau)$ правильно понимать как штраф за нарушение необходимых свойств решения, которые устанавливаются на основе физических соображений, эмпирической или теоретической информации об общих свойствах измеряемой величины. Поэтому $S(\tau)$ выбирается таким образом, чтобы её максимальное значение достигалось для распределений τ , обладающих заданными свойствами. Например, если для предполагаемого непрерывного распределения характерно отсутствие больших флуктуаций при переходе к соседнему бину, то естественно выбрать соответствующую регуляризационную функцию, которая принимает большие значения при малых отклонениях распределения в соседних бинах. Примером такой функции, соответствующей свойству гладкости распределения, может служить

$$S_1(\tau) = - \sum_{i=1}^{n-1} (\tau_{i+1} - \tau_i)^2$$

(аналог производной первого порядка). При использовании разности второго порядка регуляризационное слагаемое имеет вид

$$S_2(\tau) = - \sum_{i=2}^{n-1} (\tau_{i+1} - 2\tau_i + \tau_{i-1})^2,$$

что соответствует регуляризации Тихонова. Аналогично можно рассматривать функции, отражающие гладкость более высокого порядка. Также можно учитывать и неравномерность разбиения на бины, что, с одной стороны, усложняет вид регуляризационного слагаемого, а с другой — повышает адекватность оценки флуктуаций распределения. Отдельные методы используют регуляризационные слагаемые, соответствующие энтропии искомого распределения, например

$$S_e(\tau) = - \sum_{i=1}^n \tau_i \ln \tau_i.$$

Отметим, что возможно использование комбинации нескольких регуляризационных слагаемых.

При $\alpha = 0$ решение задачи максимизации совпадает с ранее полученным прямым решением задачи (со всеми присущими ему недостатками). При больших значениях α получается распределение, обладающее заданными свойствами, но практически не зависящее от выборки (модельно-независимое).

Методы, основанные на идее регуляризации, различаются следующими особенностями:

1) выбором вида функции правдоподобия и функции регуляризации:

- метод SVD unfolding (Картвелишвили, Хёккер [36, 44]) — регуляризационная функция, отражающая непрерывность распределения;

- метод TUnfold из пакета ROOT (Шмитт [16, 37, 53]) — комбинация из двух регуляризационных функций, отражающих непрерывность или гладкость распределения, а также ограничивающих смещение оценки;

- методы, основанные на энтропии: метод MRX (Method of Reduced Cross-Entropy) (Шмеллинг [54]), метод восстановления изображений (Нараян, Нитьянан-

да [55]), метод ARU (Automatic Regularized Unfolding) (Дембинский, Рот [56]);

2) выбором параметра регуляризации, для которого будет достигнут необходимый баланс между смещённостью оценки распределения и разбросом (статистической ошибкой) при построении точечной оценки распределения.

3.2.1. Метод SVD unfolding. Этот метод основан на регуляризации Тихонова и решении соответствующей системы уравнений методом сингулярного разложения матриц, позволяющим находить приближённые решения некорректно поставленных (например, переопределённых) систем линейных уравнений [36, 44]. Задача сводится к минимизации следующей функции, основанной на ранее описанной функции правдоподобия:

$$\Phi(\tau) = (R\tau - m)^T(R\tau - m) + \alpha \sum_{i=2}^{n-1} (\tau_{i-1} - 2\tau_i + \tau_{i+1})^2.$$

Кратко опишем данный подход. Сингулярным разложением матрицы A называется её представление в следующей форме:

$$A = USV^T,$$

где действительная матрица A имеет размер $l \times k$, U и V — ортогональные матрицы размером $l \times l$ и $k \times k$ соответственно, S — диагональная матрица $l \times k$ (диагональные значения s_{ii} неотрицательны и называются сингулярными значениями, а матрица S называется сингулярной матрицей). Выбором матриц U и V можно добиться того, чтобы сингулярные значения шли в порядке убывания. В случае $k = l$ матрицы U , S , V — квадратные ($k \times k$).

Точное решение системы линейных уравнений $A\tau = m$ с помощью SVD выполняется следующим образом. Система записывается в виде

$$USV^T\tau = m.$$

Сделаем замену переменных:

$$z = V^T\tau, \quad d = U^Tm.$$

В результате система приобретает вид

$$Sz = d,$$

её решение записывается как

$$z_i = \frac{d_i}{s_{ii}}.$$

Это позволяет получить решение исходной системы уравнений:

$$\tau = Vz.$$

Для плохо обусловленных матриц A характерна следующая картина: при решении системы описанным способом значения d_i и соответствующие s_{ii} малы, поэтому их малые возмущения приводят к сильным погрешностям решений z_i и соответствующих решений исходной системы τ .

Прямой метод решения задачи восстановления спектра представляет собой явное решение системы $R\tau = m$, эквивалентное решению задачи минимизации:

$$(R\tau - m)^T(R\tau - m) \rightarrow \min_{\tau}.$$

Авторы метода обратной свёртки на основе SVD предлагают использовать подход к нахождению истинного спектра с помощью регуляризации Тихонова, что позволяет поставить соответствующую задачу минимизации:

$$\Phi(\tau) = (R\tau - m)^T(R\tau - m) + \alpha(C\tau)^T(C\tau) \rightarrow \min_{\tau},$$

где α — коэффициент регуляризации (будет определён ниже), а матрица C удовлетворяет равенству

$$(C\tau)^T(C\tau) = \sum_{i=2}^{n-1} (\tau_{i-1} - 2\tau_i + \tau_{i+1})^2.$$

Решение описанной задачи минимизации получается как решение переопределённой расширенной системы уравнений:

$$\begin{bmatrix} AC^{-1} \\ \sqrt{\alpha}I \end{bmatrix} C\tau = \begin{bmatrix} m \\ 0 \end{bmatrix}.$$

Отметим, что матрица C вырожденная, поэтому у неё не существует обратной матрицы. Для решения этой проблемы рекомендуется рассмотреть модифицированную матрицу $\tilde{C} = C + \xi I$, где параметр ξ выбирается так, чтобы построение обратной матрицы стало возможным.

Для уменьшения погрешностей авторы метода [36, 44] рекомендуют также провести перенормировку [36] уравнений исходной линейной системы $R\tau = m$ (чтобы каждое уравнение вносило одинаковый "вклад"), при которой ковариационная матрица, соответствующая правой части исходной системы (измеренным значениям), была бы единичной. Достигается это следующим алгоритмом:

1. Правая часть исходной системы представляется в матричном виде, в результате система записывается как $R\tau = B$. Для текущей реализации $B = \text{diag } m$, в общем виде B — ковариационная матрица вектора измеренных значений.

2. Правая часть полученной системы разлагается следующим образом (сингулярное разложение):

$$B = Q\Lambda Q^T,$$

где $l_i^2 = \Lambda_{ii} > 0$, $\Lambda_{ij} = 0$ для $i \neq j$,

$$B^{-1} = Q\Lambda^{-1}Q^T.$$

3. Левая и правая части уравнения нормируются:

$$\tilde{R}_{ij} = \frac{1}{l_i} \sum_k Q_{ik} R_{kj}, \quad \tilde{m}_i = \frac{1}{l_i} \sum_k Q_{ik} m_k.$$

В результате получается перенормированная система уравнений: $\tilde{R}\tau = \tilde{m}$, к которой применяются все описанные ранее операции (введение регуляризационного слагаемого и решение расширенной системы). Решение новой системы проводится описанным выше способом, однако промежуточные решения z_i теперь выглядят так:

$$z_i = \frac{d_i}{s_{ii}} \frac{s_{ii}^2}{s_{ii}^2 + \alpha}.$$

Отметим, что в случае $s_{ii}^2 \gg \alpha$ решения практически совпадают с рассмотренным прямым решением исходной задачи. При малых сингулярных значениях s_{ii} у же

становится более заметным влияние дополнительного регуляризационного слагаемого. Для нахождения "эффективного значения" параметра регуляризации рекомендуется визуально проанализировать сингулярные значения в логарифмической шкале: построить график $\log s_{ii}$, выделить его линейный участок и выбрать значение $i = k$, при котором монотонное линейное убывание $\log s_{ii}$ сменяется резким падением. В качестве коэффициента регуляризации выбирается $\alpha = s_{ii}^2$ ("последнее большое" сингулярное значение s_{kk}) [36]. Отметим, что возможно использование описанного алгоритма восстановления спектра и при других значениях коэффициента регуляризации.

3.2.2. Метод TUnfold. На идее регуляризации также основан алгоритм, реализованный в пакетах ROOT и RooUnfold в виде процедуры TUnfold [16, 37, 53]. Он предполагает минимизацию функции правдоподобия следующего вида:

$$\Phi(\tau) = L_1(\tau) + \alpha L_2(\tau) + \lambda L_3(\tau).$$

Здесь слагаемое $L_1 = (R\tau - m)^T(R\tau - m)$ соответствует исходной функции правдоподобия в прямом методе восстановления спектра: если коэффициенты при остальных слагаемых обнулить, то оценка истинного распределения как точка минимума функции правдоподобия будет совпадать с построенной на основе прямого метода, что на практике приводит к большим погрешностям. Слагаемое L_2 обеспечивает соответствие предполагаемым характеристикам исходного распределения (например, непрерывности или гладкости). В качестве такого слагаемого могут быть выбраны функции S_1 или S_2 , ранее введенные в общем описании методов, основанных на регуляризации. Наконец, в последнем слагаемом используется функция $L_3(\tau) = \sum_i m_i - e^T \tau$, где компоненты вектора e задаются как $e_j = \sum_i R_{ij}$. Это слагаемое необходимо для ограничения излишне большого смещения получаемой оценки распределения [37].

В методах, использующих идею регуляризации, важным этапом решения задачи является способ выбора коэффициентов при регуляризационных слагаемых. Например, в описанном в разделе 3.2.1 SVD-методе коэффициент регуляризации выбирался на основе анализа сингулярных значений, однако такой способ сильно зависит от самого метода решения задачи восстановления распределения. В алгоритме TUnfold [37] коэффициенты регуляризации не задаются, а автоматически выбирают-ся одним из двух методов:

1. *Метод L-curve.* В основе метода лежит решение задачи максимизации функции $\Phi(\tau) = \ln L(\tau) + \alpha S(\tau) \rightarrow \max_\tau$. Такая задача последовательно решается для разных значений параметра α в некотором диапазоне с некоторым шагом. Значения слагаемых $L(\tau)$ и $S(\tau)$, для которых при текущем значении параметра α целевая функция достигает максимума, обозначим как L_α и S_α . Прологарифмируем эти значения: $\mathcal{L}_x(\alpha) = \log L_\alpha$, $\mathcal{L}_y(\alpha) = \log S_\alpha$, после чего построим параметрическую кривую по точкам $(\mathcal{L}_x(\alpha), \mathcal{L}_y(\alpha))$ (для обеспечения должной гладкости построенная кривая аппроксимируется кубическим сплайном). В качестве искомого значения параметра α для регуляризационного слагаемого выбирается тот, для которого кривизна построенной кривой в соответствующей точке максимальна [12, 13, 37].

2. *Метод минимизации коэффициентов корреляции.* Значения τ_i можно рассматривать как набор случайных величин, для которых оценивается ковариационная матрица $V_{\tau\tau}$. Для построения этой оценки при выбранном значении коэффициента регуляризации α проводится серия вычислительных экспериментов по восстановлению спектра, в результате формируется выборка $(\tau_1^k, \dots, \tau_n^k)$, где $k = 1, \dots, K$. По данной выборке строится оценка ковариационной матрицы $V_{\tau\tau}$. Значения глобального коэффициента корреляции ρ_i для i -го бина определяются при разных значениях α следующим образом:

$$\rho_i(\alpha) = \sqrt{1 - \frac{1}{(V_{\tau\tau}^{-1})_{ii}(V_{\tau\tau})_{ii}}}.$$

В качестве оптимального значения коэффициента регуляризации выбирается такое α , при котором достигается минимальная величина среднего арифметического $\rho_i(\alpha)$ или максимального значения из всех $\rho_i(\alpha)$ [2, 12, 37].

Отметим, что эти методы выбора параметров регуляризации, предложенные для TUnfold, являются универсальными и могут быть применены в других алгоритмах. Для той же задачи используются и другие подходы, среди которых можно выделить метод предельной оценки максимального правдоподобия (Modified Maximum Likelihood Estimator, MMLE) [41], метод обобщенной (взвешенной) кроссвалидации [57], информационный критерий Акаике [58, 59] и остановку по критерию согласия [60].

Идея применения регуляризации Тихонова для обратной свертки используется достаточно давно. В физике элементарных частиц она начала широко применяться после разработки регуляризационного алгоритма, реализованного в пакете *RUW* [1, 61].

Существуют и другие методы решения задачи обратной свертки, основанные на той же идее. В частности, можно отметить подходы, в которых вид слагаемого регуляризации основан на энтропии распределения случайной величины (регуляризация по методу максимальной энтропии) [55, 62]. В этом случае оно может иметь вид

$$S_e(\tau) = - \sum_{i=1}^n \tau_i \log \tau_i \quad \text{или} \quad S_{\text{ent}}(\tau) = - \sum_{i=1}^n \frac{\tau_i}{N} \log \frac{\tau_i}{N}.$$

Во втором случае в качестве коэффициента регуляризации авторы метода предлагают использовать $\alpha = 1/N$, однако на практике это приводит к излишне сглаженному распределению [11] (максимум энтропии соответствует дискретному равномерному распределению), поэтому рекомендуется использовать меньшие значения регуляризационного коэффициента.

Кроме того, в работе [63] предложен метод, основанный на минимизации дисперсии истинного распределения и использовании идеи комбинирования регуляризационных слагаемых разного вида. Выбор конкретного метода регуляризации, как правило, обусловлен априорной информацией, характеризующей особенности исходного распределения и его искажения в ходе измерения.

Идея применения SVD-подхода также была развита в методе Wiener-SVD [64]. Помимо описанной ранее техники SVD-алгоритма авторы дополнительно применяют фильтр Винера, позволяющий уменьшить влияние шу-

мов. В то же время построение фильтра Винера требует достаточно хорошей оценки истинного спектра, что затрудняет применение алгоритма.

Подход, основанный на регуляризации, применяется и в итеративных алгоритмах обратной свёртки. В частности, метод IDS (Iterative, Dynamically Stabilized Method of Data Unfolding) [65, 66] использует динамически изменяющееся регуляризационное слагаемое, позволяя тем самым уменьшить флуктуации восстанавливаемого спектра на соседних итерациях.

3.3. Метод поправочных коэффициентов

Помимо описанных в разделах 3.1, 3.2, существуют и другие подходы к восстановлению истинного спектра. Например, одним из наиболее простых является так называемый метод поправочных коэффициентов (метод bin-by-bin correction). В простой его реализации в ходе численного моделирования получают истинное и измеренное (искажённое) распределения рассматриваемой физической величины. Далее для каждого интервала вычисляется поправочный коэффициент C_i , являющийся оценкой того, во сколько раз измеренное количество частиц, попавших в интервал, отличается от истинного. Оценка истинного спектра вычисляется делением измеренного спектра на поправочный коэффициент в каждом из интервалов: $\hat{\tau}_i = \mu_i / C_i$.

В описанной форме рассматриваемый подход имеет ряд недостатков, ограничивающих его использование на практике [12, 67, 68]. Корректный расчёт поправочных коэффициентов предполагает, что уже известно истинное распределение частиц по бинам, но именно оценка этого распределения и является основной задачей. Здесь стоит отметить, что процедура вычисления поправочных коэффициентов отличается от нахождения матрицы миграций: в первом случае ключевым является определение истинного распределения или его хорошей оценки, а во второй задаче это не играет столь важной роли — достаточно оценивать вероятности, с которыми частица детектируется в других бинах (причём для построения оценки важно иметь достаточно большое количество событий, но информация о том, как эти события должны быть распределены по интервалам, не принципиальна). Наконец, описанный простой подход не учитывает миграцию между бинами (даже соседними), поэтому в реализации не предполагается обратного перераспределения событий между интервалами, привлекается лишь механистическая корректировка (в результате общее количество событий может измениться).

Также существуют модификации метода поправочных коэффициентов, в определённой степени сглаживающие отмеченные недостатки. Например, для восстановления спектра можно использовать итеративный процесс, в котором описанная выше процедура вычисления поправочных коэффициентов является только первым шагом [69, 70]. Этот подход предполагает уточнение поправочных коэффициентов на очередном шаге алгоритма с использованием предыдущей оценки спектра и корректировку параметров метода Монте-Карло, моделирующего приборные искажения спектра [69].

Тем не менее применение метода поправочных коэффициентов (в том числе в простой форме) является допустимым, если миграция между бинами невелика (матрица миграций близка к диагональной) [8, 11, 32]. При соблюдении этого условия метод используется при об-

работке экспериментальных данных в современной физике частиц [43, 71, 72].

3.4. Прочие методы восстановления спектра

Наряду с перечисленными существуют и другие методы обратной свёртки, использующие как уже отмеченные подходы (итерационные методы построения оценок, регуляризацию, побиноую корректировку распределений), так и другие математические идеи и принципы. В частности, алгоритмы, основанные на итеративном построении статистических оценок распределения неизвестной величины, используются достаточно давно. В начале 1930-х годов ван Ситтерт предложил итеративный метод восстановления искажённых спектров [73, 74], в котором текущая оценка спектра корректируется на величину, зависящую от того, насколько искажение текущей оценки отклоняется от результатов измерения:

$$\tau^{(k+1)} = \tau^{(k)} + \alpha(m - R\tau^{(k)}).$$

Развитием этого подхода является предложенный в 1964 г. алгоритм Голда [75], в дальнейшем применяемый в физике частиц [18, 21, 76]. Ключевое отличие такого метода заключается в использовании для различных бинов соответствующих коэффициентов [75, 77]:

$$\alpha_i^{(k)} = \frac{\tau_i^{(k)}}{\sum_{j=1}^n R_{ij}\tau_j^{(k)}},$$

а оценка спектра на очередной итерации строится как

$$\tau_i^{(k+1)} = \tau_i^{(k)} + \alpha_i \left(m_i - \sum_{j=1}^n R_{ij}\tau_j^{(k)} \right).$$

Итеративную процедуру алгоритма Голда можно также записать как последовательное уточнение текущей оценки величины в текущем бине посредством домножения на некоторый поправочный коэффициент. Схожий подход используется в упомянутом ранее алгоритме Ричардсона – Люси [45, 46], изначально созданном для обработки астрономических изображений [78] и в дальнейшем адаптированном в том числе для задач восстановления спектров в экспериментах по физике частиц [79, 80].

Некоторые алгоритмы обратной свёртки разрабатываются для решения узкоспециализированной физической задачи или анализа особенностей конкретных измерительных приборов. Так, при изучении потоков нейтронов используются методы обратной свёртки, специфические для решаемой задачи восстановления спектров, полученных с помощью сфер Боннера. Например, среди таких методов можно упомянуть алгоритмы SPUNIT [81], BON [82], SAND-II [83, 84], MAXED [85], GRAVEL [86].

Описанные ранее методы обратной свёртки использовались для восстановления распределений физической величины с предварительным разбиением рабочего диапазона её значений на бины. Полученные оценки являлись дискретизированными оценками истинного распределения. Однако в таком подходе отмечают некоторые недостатки или особенности, затрудняющие его использование на практике [24]. В первую очередь это необходимость заранее разбивать рабочий диапазон на интервалы, причём такая процедура выполняется в большин-

стве случаев вручную [61, 87–90]. Во-вторых, группирование данных по бинам является огрублением [61], не позволяющим гибко учесть все дополнительные эффекты, влияющие на отклик прибора, что приводит к неоптимальным оценкам спектра. Наконец, гистограммные методы трудно адаптируются при восстановлении распределения многомерной случайной величины. Отмеченные особенности заставляют обратить внимание на подходы, в которых оценка истинного спектра выполняется без предварительного разбиения диапазона значений на интервалы.

Для построения оценок плотности распределения без дискретизации могут быть использованы подходы, основанные на применении методов машинного обучения [91–94], которые также пригодны и для оценки дискретизированных распределений. Их основная идея заключается в использовании результатов численного моделирования в качестве обучающей выборки для построения системы прогнозирования (например, модельного истинного распределения по модельному измеренному) или классификатора. В роли системы для решения задачи обратной свёртки может быть выбрана нейронная сеть [92] или регрессионная модель [93], на вход которых подаются измеренные значения рассматриваемой величины или полностью измеренный спектр, а на выходе ожидается истинное значение рассматриваемой величины или истинный номер бина, в который попадает эта величина, либо восстановленный истинный спектр.

Также используется подход, основанный на модели генеративно-сопоставительной нейронной сети (Generative Adversarial Network, GAN) [95]: генеративная сеть итеративно формирует образцы, соответствующие предполагаемому истинному распределению рассматриваемой величины, а дискриминативная сеть оценивает их с точки зрения соответствия измеренному распределению при известном характере искажений (процедура аналогична построению матрицы миграций на основе численного моделирования) [91, 96]. Затруднения при применении описанных подходов на практике аналогичны отмеченным выше проблемам, связанным с использованием поинтервальной коррекции поправочными коэффициентами.

Другим возможным подходом к описанию оцениваемой плотности распределения является её представление в виде сплайн-функции. Такой способ реализован, в частности, в алгоритме ARU [23, 56, 97], на разработку которого в значительной степени повлияли идеи, предложенные в работах [1, 54]. Здесь искомая непрерывная плотность распределения $f(x)$ представляется в виде кубического В-сплайна [98] $f(x) = \sum c_j b_j(x)$ с неизвестными (оцениваемыми) коэффициентами c_j и произвольно выбираемыми узлами. Интервалы между узлами должны быть достаточно узкими для того, чтобы адекватно отразить все особенности искомой плотности распределения, их количество и расположение не являются критичными. Искажение этой плотности распределения описывается свёрткой

$$g(y) = \int K(y, x) f(x) dx = \sum_j c_j g_j(y),$$

где искажённые базисные функции

$$g_j(y) = \int K(y, x) b_j(x) dx$$

численно оцениваются в предложенном алгоритме. Ядро свёртки $K(y, x)$ оценивается при калибровке или моделировании отклика прибора методами Монте-Карло аналогично тому, как это проводится при построении матрицы миграций в описанных ранее методах. Оценка коэффициентов c_j соответствует минимуму отрицательной функции правдоподобия $L(c) = L_1(c) + \lambda L_2(c)$. Первое слагаемое имеет вид

$$L_1(c) = \sum_j c_j G_j - \sum \ln g(y_i),$$

где $G_j = \int g_j(y) dy$, y_i — измеренные значения физической величины. Второе слагаемое представляет собой взятую с коэффициентом λ регуляризационную функцию

$$L_2(c) = \int f(x) \ln \frac{f(x)}{h(x)} dx - \sum_j c_j F_j,$$

где первая составляющая является расстоянием Кульбака–Лейблера от искомой плотности распределения f до некоторой предполагаемой плотности h , а $F_j = \int f_j(x) dx$. Выбору предполагаемой плотности $h(x)$ посвящён отдельный параграф статьи [56], авторы которой рекомендуют в качестве $h(x)$ выбирать функцию, наиболее близкую к искомой плотности распределения, если такая информация существует априорно. В качестве грубого приближения для $h(x)$ можно взять равномерное распределение на рассматриваемом диапазоне значений либо провести первое восстановление спектра с использованием равномерного распределения $h_0(x)$ (в результате получим оценку восстановленного спектра $f_0(x)$), а затем в качестве $h(x)$ использовать свёртку $f_0(x)$ с ядром $K(y, x)$ и на основе полученной функции выполнять окончательную процедуру восстановления спектра.

Некоторые подходы, применяемые в гистограммных методах обратной свёртки, могут быть адаптированы для непрерывного (безбинового) случая. Так, итеративное построение дискретизированной оценки спектра послужило основной идеей разработки алгоритма OmniFold [24].

Как отмечено ранее, итеративный байесовский алгоритм восстановления спектра основан на последовательном уточнении оценки спектра в отдельных бинах. В алгоритме OmniFold предложена идея распространить такой подход на оценку плотности распределения без разбиения на бины. Ключевым является вычисление оценки отношения правдоподобия

$$L_{(w, X), (w', X')}(x) = \frac{p_{(w, X)}(x)}{p_{(w', X')}(x)},$$

где под $p_{(w, X)}(x)$ понимается оценка плотности распределения случайной величины, полученная по выборке X с параметрами (весами) w . Эта оценка строится с помощью выбранной модели машинного обучения, в частности, авторы используют нейросетевой алгоритм PFN (Particle Flow Networks) [99, 100].

Основным подходом, применяемым в OmniFold, является рассмотрение событий на уровне детектора и уровне частиц, симулируемых методами Монте-Карло. Для обучения (подбора параметров) системы используется набор пар (t, m) , в которых событию на уровне частиц t соответствует событие на уровне детектора m , полученное (сгенерированное) симуляцией работы соответствующего детектора. Сгенерированные детекторные

события (обозначим их как Sim) преобразуются (с некоторыми параметрами) для их соответствия экспериментальным данным (Data). Преобразованные данные используются для уточнения оценки параметров и значений величин на уровне частиц (Gen), которые снова передаются на уровень детектора на следующем шаге итеративной процедуры.

Рассматриваемый итеративный алгоритм формально можно описать следующим образом.

Инициализация. Задаются начальные значения параметров (весов) $v_0(t)$ на уровне частиц. При симуляции работы детектора, преобразующего события t в m , они становятся параметрами на уровне детектора: $v_0^{\text{push}}(m) = v_0(t)$.

Далее итеративно выполняются следующие два шага.

Шаг 1. Оценка параметров на уровне детектора:

$$\omega_n(m) = v_{n-1}^{\text{push}}(m) L_{(1, \text{Data}), (v_{n-1}^{\text{push}}, \text{Sim})}(m).$$

Затем эти параметры передаются на уровень частиц: $\omega_n^{\text{pull}}(t) = \omega_n(m)$.

Шаг 2. Оценка параметров на уровне частиц:

$$v_n(t) = v_{n-1}(t) L_{(\omega_n^{\text{pull}}, \text{Gen}), (v_{n-1}, \text{Gen})}(t).$$

После выполнения n таких итераций восстановленное распределение величины вычисляется как $p_{\text{unf}}^{(n)}(t) = v_n(t) p_{\text{Gen}}(t)$. Результатом работы алгоритма может быть как набор сгенерированных событий t с весами $v_n(t)$, так и просто сгенерированная функция $v_n(t)$, которую можно применить к измеренному распределению плотностью, оцениваемой по произвольному набору данных.

Помимо основного алгоритма авторами [24] предложены его версии, использующие более простую трёхслойную модель нейронной сети для генерирования функции $L_{(w, X), (w', X')}(x)$. Эти версии могут быть применены как для восстановления одномерных распределений (алгоритм UniFold), так и в многомерном случае (алгоритм MultiFold). Также авторы [24] отмечают, что предложенный метод может быть использован и в бинарном случае (для построения гистограммных оценок плотности) и адаптирован для других схем генерирования отношения правдоподобия [101, 102].

3.5. Программные реализации

Прикладная важность задачи обратной свёртки, большие объёмы подлежащих обработке данных и распространённость областей применения соответствующих методов требуют разработки программных реализаций описанных в разделах 3.1–3.4 алгоритмов.

Первым распространённым программным продуктом для решения таких задач в физике элементарных частиц стал разработанный на языке FORTRAN 77 пакет *RUN* [1, 61], в котором был реализован регуляризационный алгоритм обратной свёртки, а также сглаживание распределений с помощью В-сплайнов [2].

В настоящее время для обработки экспериментальных данных в физике элементарных частиц активно используется программный пакет ROOT [49], разработанный в ЦЕРНе. В частности, в ROOT реализован алгоритм TUnfold [37] на основе идеи регуляризации. Помимо собственно алгоритма обратной свёртки в пакете имеются средства разбиения данных на бины, а кроме

того, реализована возможность объединения данных для анализа из различных источников.

Также в среде ROOT разработан пакет RooUnfold [50], особенностью которого является разнообразие реализованных в нём алгоритмов обратной свёртки: итеративный байесовский алгоритм д'Агостини [38, 40], метод SVD-Unfold [36], интерфейс к реализованному в ROOT алгоритму TUnfold, алгоритм IDS [66], метод поправочных коэффициентов. В дальнейшем эти же методы были включены в программный пакет RooFitUnfold [103]. На основе библиотек ROOT также был разработан пакет прикладных программ TRUEE (Time-dependent Regularized Unfolding for Economics and Engineering problems) [104], являющийся версией пакета *RUN* на C++, адаптированной для более широкого использования.

Основная масса программных разработок представляет собой прикладную реализацию одного алгоритма или некоторого класса методов. Например, пакет PyUnfold [105] содержит реализацию итеративного байесовского алгоритма [40], модифицированного за счёт более широкого набора метрик, используемых для остановки итеративной процедуры. Алгоритм на основе SVD-подхода [36] изначально был реализован в программном пакете GURU [44]. К этой же категории программных разработок относятся упомянутые ранее пакеты *RUN* и TRUEE, реализующие один из регуляризационных методов и широко применявшиеся до появления более эффективных алгоритмов в ROOT и RooUnfold. Алгоритм OmniFold использует разработанный авторами [106] пакет EnergyFlow, включающий программные инструменты для задач физики частиц, и также является доступным с открытым кодом на Python [107].

В некоторых коллаборациях создаются собственные реализации методов, основанные, однако, на всё тех же подходах. Например, в коллаборации PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics) существует пакет программ PamUnfold, в котором реализованы обе версии (базовая и модифицированная) байесовского алгоритма д'Агостини. В отдельных областях используются узкоспециализированные программные реализации алгоритмов, изначально спроектированные для решения конкретных экспериментальных задач. Например, упомянутые ранее алгоритмы восстановления спектров нейтронов SPUNIT и BON реализованы в программном пакете BUNKI [108] и его модификации [109], для алгоритма MAXED используется одноимённый программный модуль. Также эти алгоритмы и метод SAND-II программно реализованы в пакете BUMS [110], особенностью которого является наличие веб-интерфейса, позволяющего получить доступ к программе через браузер. В то же время созданные для решения локальных задач методы и реализующие их программные пакеты также могут получить и широкое распространение. Так, реализация алгоритма *sPlot* [112] для обработки экспериментальных результатов коллаборации BaBar [111] включена во фреймворк ROOT (класс TSPlot) и в настоящее время используется для решения разнообразных задач физики частиц [113–116].

4. Использование методов обратной свёртки в экспериментах по физике частиц

Эксперименты в различных областях физики элементарных частиц оказываются на данный момент главными,

где активно применяются различные методы обратной свёртки. Наиболее ярким примером их использования и основным драйвером развития являются эксперименты на ускорителях, где часто можно встретить разные подходы и алгоритмы, являющиеся неотъемлемой частью анализа научных данных.

Во многих публикациях описываются методы обратной свёртки, применяемые к решению задач в экспериментах на ускорителях DESY (нем. Deutsches Elektronen-Synchrotron) [117], Тэватрон [118–121], SPS (Super Proton Synchrotron) [122] и LHC [1, 123]. В качестве примеров можно привести работы с анализом данных детекторов ATLAS (A Toroidal LHC ApparatuS) [124, 125] и CMS (Compact Muon Solenoid) [126, 127]. Существуют подробные обзоры, посвящённые применению методов обратной свёртки в указанных экспериментах [41, 128].

Методы обратной свёртки так или иначе применяются в анализе большинства изучаемых на ускорителях явлений. Особый интерес в последнее время представляет исследование процессов, связанных с распадами бозона Хиггса и реакциями с участием t -кварков [129, 130]. Это обусловлено тем, что информация об истинных распределениях физических величин в данных LHC важна при поиске сигналов от так называемой Новой физики — гипотетических процессов, не укладывающихся в рамки Стандартной модели. Поскольку эффекты от таких процессов ожидаются минимальными, необходимо устранение даже самых незначительных приборных искажений. Например, в работе [131] моделируется экзотический распад бозона Хиггса с генерацией адронов в конце цепочки распадов. Сигнал от такого процесса должен быть виден как избыток ("бамп") на распределении инвариантной массы системы Z -бозона и адронной струи, к которому и применяется алгоритм обратной свёртки (в данном случае OmniFold).

Нейронные сети на основе модели генеративно-состязательных сетей используются в анализе приборных эффектов детекторов на ускорителе LHC, возникающих при изучении рождения пары Z - и W -бозонов и процессов, описываемых квантовой хромодинамикой [132]. Подробное исследование применимости методов обратной свёртки к данным детектора ATLAS на LHC проведено в диссертации [133].

Предполагается использование методов обратной свёртки и в экспериментах на перспективных ускорительных комплексах, например, при обработке данных детектора MPD (Multi-Purpose Detector), который будет размещён на строящемся в Дубне ускорителе NICA (Nuclotron-based Ion Collider fAcility) [134–136]. В данном случае методы анфолдинга применяются уже на этапе численного моделирования характеристик детектора MPD с целью получить максимально детальное и приближённое к реальности описание его характеристик [137, 138].

Другим важным научным направлением, где часто применяются методы обратной свёртки при анализе данных, является физика космических лучей. Здесь следует рассмотреть две категории экспериментов: в космическом пространстве (на космических аппаратах и МКС) и на наземных установках.

Нахождение в космическом пространстве накладывает большое количество ограничений на исполнение прибора (физические размеры, потребляемую мощность и др.), что сказывается как на энергетическом разреше-

нии приборов, так и на эффективности регистрации частиц, которая может зависеть от разных факторов. Это может, например, привести к неправильному измерению показателя спектра. В современных экспериментах, основанных на магнитной спектроскопии, таких как BESS (Balloon-borne Experiment with Superconducting Spectrometer), PAMELA и AMS-02 (Alpha Magnetic Spectrometer-02), методы обратной свёртки применялись при анализе опубликованных энергетических спектров космических лучей (см., например, [139–146]).

При энергиях частиц свыше нескольких десятков ТэВ измерения, основанные на магнитном анализе, становятся невозможными. Обеспечить измерения космических лучей с энергиями до нескольких ПэВ способны приборы, определяющие характеристики частиц калориметрическими методами. Для них в космосе основным фактором является размер калориметра, а энергия измеряется по суммарному энерговыделению в каскаде, вызванном частицей космических лучей. При этом с возрастанием энергии первичной частицы пространственные размеры каскада перестают укладываться в размерах калориметра. Последнее приводит к ухудшению разрешающей способности прибора и возрастанию ошибки при измерении энергии. Как и в предыдущем случае, для более надёжного определения показателя спектра необходимо использовать методы обратной свёртки. Примеры их применения в подобных измерениях приведены в работах [72, 147] для экспериментов DAMPE (DARK Matter Particle Explorer) и CALET (CALorimetric Electron Telescope) соответственно.

С дальнейшим возрастанием энергии регистрация частиц в экспериментах в космосе становится невозможной, и для изучения характеристик космических лучей сверх- и ультравысоких энергий используются наземные установки, принцип работы которых основан на регистрации широких атмосферных ливней (ШАЛ). Существуют две независимые техники регистрации ШАЛ: с помощью массива так называемых наземных детекторов (SD), регистрирующих заряженную компоненту ШАЛ, достигающую поверхности Земли, и с использованием оптических систем, наблюдающих флуоресценцию при развитии каскада в атмосфере (FD). Техники могут сочетаться, что повышает точность получаемых данных. Тем не менее искажение измеряемых характеристик может оказаться значительным из-за таких факторов, как, например, засветка от других природных явлений, атмосферных эффектов, больших размеров ШАЛ частиц ультравысокой энергии и необходимости размещать детекторы на расстоянии друг от друга, чтобы покрывать большую площадь. Как и при калориметрических измерениях, не удаётся с одинаковой эффективностью регистрировать ШАЛ от точки рождения до его полного поглощения.

Примеры использования методов обратной свёртки в подобных случаях можно встретить в публикациях по результатам измерения спектров космических лучей ультравысоких энергий такими обсерваториями, как Обсерватория им. Пьера Оже (Pierre Auger Observatory, PAO) [148] или Telescope Array [149].

Схожие методы космических и наземных измерений применяются при регистрации гамма-излучения в гамма-астрономии. При этом сохраняются и все упомянутые выше искажения, что также требует их коррекции. В работах [150, 151] приведено подробное описание методов обратной свёртки, которые применяются в экс-

перименте MAGIC (Major Atmospheric Gamma Imaging Cherenkov telescopes), а в [152] — в эксперименте H.E.S.S. (High Energy Stereoscopic System). Большой обзор таких методов в гамма-спектрометрии можно найти в [153]. В работах [154–156] рассказывается о применении методов обратной свёртки при обработке энергетических спектров точечных источников гамма-излучения, а в работе [157] они применены к суммарному дифференциальному энергетическому спектру электронов и позитронов, измеренному космической гамма-обсерваторией Fermi.

Широко применяются методы восстановления спектра и в нейтринной физике. Нейтрино из-за чрезвычайно малого сечения взаимодействия с веществом очень сложно зарегистрировать. Требуются значительные объёмы вещества детектора, при прохождении через которое нейтрино высокой энергии могут вступать во взаимодействие с ядрами, образуя вторичные заряженные частицы, например электрон или мюон. Вторичные частицы регистрируются детекторами, определяются их характеристики, по которым далее восстанавливается информация об энергии и направлении прилёта нейтрино. Размеры установки, необходимость располагать детекторы на большом расстоянии друг от друга, качество вещества мишеней (особенно, если ими являются природные резервуары льда или воды), а также другие факторы приводят к ухудшению разрешающей способности установки. Таким образом, учёт приборных эффектов становится крайне важной задачей при восстановлении энергетических спектров и пространственных распределений нейтрино различной природы.

Так, в эксперименте IceCube на Южном полюсе для восстановления истинного спектра атмосферных мюонных нейтрино процедура восстановления спектра применяется к распределениям количества зарегистрированных мюонов и их энергетических потерь [158, 159]. Подобные методы применяются в экспериментах ANTARES (Astronomy with a Neutrino Telescope and Abyss environmental REsearch) в Адриатическом море и Baikal-GVD (Baikal Gigaton Volume Detector) на озере Байкал [160].

Кроме научных задач методы обратной свёртки применяются и в практических целях: измерение светимости

адронного колайдера через упругие рассеяния протонов друг на друге, где восстанавливается истинное распределение по углу рассеяния [161]; задачи нейтронной спектроскопии [162, 163]; изучение спектра реакторных антинейтрино [164]; мониторинг, обнаружение и идентификация радиоактивных материалов по форме спектра ионизирующего излучения [153] и др.

Важно отметить, что восстановление истинных распределений требуется не только при анализе конечных результатов, таких как спектры частиц. С применением методов обратной свёртки восстанавливаются амплитуды сигналов фотоэлектронных умножителей [19, 165–167], распределения сигналов в кластерах позиционно-чувствительных детекторов, например, на установках ALICE/LHC [168] или MPD/NICA [134–138], и решаются другие подобные задачи [169]. Это позволяет получить более точную информацию о времени пролёта и скорости частиц, лучше восстановить их координаты и траекторию.

Методы обратной свёртки применяются в физике не только высоких, но и низких энергий, например, в нейтронной спектрометрии и гамма-спектрометрии, ядерной медицине, биомедицине и других направлениях [170–172].

В таблице приведён список некоторых наиболее известных экспериментов в области физики элементарных частиц, космофизики и гамма-астрономии, где используются методы обратной свёртки для восстановления истинных характеристик измеряемых величин. Также указаны примеры конкретных публикаций, краткая формулировка решаемых физических задач и используемые методы обратной свёртки.

5. Перспективы развития методов обратной свёртки. Заключение

Использование методов обратной свёртки для восстановления истинных распределений физических величин сегодня находит широкое применение в различных научных областях, активно обсуждается на научных конференциях и специально организованных школах (Terascale Statistics School, Pan-European Advanced

Таблица. Примеры использования методов обратной свёртки для восстановления истинных распределений измеряемых физических величин

Эксперимент	Физическая задача	Литература	Метод
Эксперименты на ускорителях			
ATLAS/LHC	Измерение сечения четырёхлептонного распада бозона Хиггса	[71]	Метод поправочных коэффициентов
CMS/LHC	Измерение зарядовой асимметрии в рождениях топ-кварковых пар в протон-протонных столкновениях	[126]	TUnfold
ALFA/LHC	Измерение сечения упругого рассеяния в столкновениях протонов для измерения светимости LHC	[161]	Итеративные методы
Тэватрон/Фермилаб	Изучение свойств рождения и распада пары топ-кварка и антикварка	[118]	Регуляризация
HERA/DESY	Изучение корреляций в рождениях лептонов и адронных струй в глубоко-неупругих рассеяниях протонов	[117]	OmniFold
KLOE/DAΦNE*	Измерение пионного формфактора в электрон-позитронных столкновениях для изучения аномального магнитного момента мюона	[173]	Байес
BEPC/BES III**	Изучение процессов рождения и распада Ψ -мезонов	[174]	Другие методы
CBM/FAIR***	Исследование возможных отклонений от закона сохранения заряда в столкновениях золото–золото	[175]	Байес

Таблица (окончание)

Эксперимент	Физическая задача	Литература	Метод
Космические лучи			
AMS-02	Магнитоспектрометрические измерения энергетических спектров протонов галактических космических лучей (ГКЛ)	[176]	Байес
FERMI	Калориметрические измерения энергетического спектра протонов ГКЛ	[177]	TUnfold
PAMELA	Магнитоспектрометрические измерения спектров антипротонов, позитронов и ядер	[140, 141, 178]	Байес
DAMPE	Калориметрические измерения спектров протонов и ядер гелия в ГКЛ	[147, 179]	Байес
CALET	Калориметрические измерения спектров ядерной компоненты ГКЛ	[72, 180]	RooUnfold
PAO	Измерение спектра космических лучей с энергиями выше $2,5 \times 10^{18}$ эВ	[181]	Метод поправочных коэффициентов
Гамма-астрономия			
MAGIC	Измерение спектра гамма-излучения от пульсаров	[182, 183]	Регуляризация, "наивный" подход
H.E.S.S.	Измерение спектров гамма-излучения от астрофизических источников в Крабовидной туманности и PKS 2155-304	[152]	Байес
Изучение нейтрино			
IceCube	Измерение потоков астрофизических нейтрино от точечных источников	[184, 185]	<i>RUN</i> (TRUEE)
ANTARES, Baikal-GVD	Измерение энергетических спектров атмосферных нейтрино	[160]	—
Супер-Камиоканде	Измерение потоков атмосферных нейтрино и антинейтрино	[186]	Пакет RooUnfold
Daya Bay	Измерение спектров реакторных антинейтрино	[164]	Wiener-SVD
* K LOnG Experiment/Double Annular Φ Factory for Nice Experiments. ** Beijing Electron – Positron Collider/Beijing Spectrometer III. *** Compressed Baryonic Matter/Facility for Antiproton and Ion Research.			

School on Statistics in High Energy Physics и др.), включается в учебные курсы по статистическим методам и их приложениям [8, 187–189]. Однако не существует общего универсального подхода, успешно используемого в разных опытах и экспериментах, и в каждом случае требуется дополнительное исследование. В связи с этим важно развивать все возможные подходы и предлагать новые методы для решения задачи обратной свёртки.

В то же время требуется подробный сравнительный анализ различных методов, позволяющий установить их ограничения и особенности использования — например, выявить зависимость результата от области определения значений физической величины, способов её разбиения на интервалы, особенностей спектра физической величины и других параметров. Также необходимо предложить способ оптимального выбора параметров самих методов и оценки погрешности при восстановлении истинного распределения разными алгоритмами [190], что позволит сделать выводы о наилучшей применимости методов к разным научным задачам.

На основе подобного анализа возможна разработка новых алгоритмов решения поставленной задачи, в том числе на основе математических подходов, ранее не применявшихся в методах обратной свёртки. Такие работы последнее время появляются в печати (см., например, [24, 32, 42, 64, 191–193]).

Другими важными задачами являются разработка новых и развитие существующих методов восстановления для многомерного распределения [67, 194, 195]. Для алгоритмов на основе байесовского подхода (итеративный алгоритм д'Агостини, методы IDS и FBU) критич-

ным является уменьшение среднего числа событий в бине (при сохранении среднего размера количество бинов резко растёт при увеличении размерности пространства). Также отметим, что применяемая в одномерном случае регуляризация Тихонова предполагает присвоение соседним бинам соседних (отличающихся на 1) номеров. В многомерном случае в общем виде не существует такой нумерации многомерных бинов, при которой пространственно близкие бины имели бы близкие номера. По этой причине явный перенос регуляризационных методов на многомерный случай затруднён и требует адаптации. Тем не менее ранее описанные прикладные задачи анализа эффективности различных алгоритмов восстановления спектра не теряют своей актуальности и сложности и в многомерном случае.

Исследование выполнено за счёт гранта Российского научного фонда (проект № 19-72-10161).

Список литературы

1. Blobel V, in *Proc. of the 1984 CERN School of Computing, Aiguablava, Catalonia, Spain, 9–22 September 1984* (CERN 85–09) (Geneva: CERN, 1985) pp. 84–114
2. Blobel V, hep-ex/0208022
3. Rust B W, Ingersoll D T, Burrus W R *A User's Manual for the FERDO and FERD Unfolding Codes* (Oak Ridge, TN: Oak Ridge National Laboratory, 1983)
4. Engl H W, Hanke M, Neubauer A *Regularization of Inverse Problems* (Dordrecht: Kluwer Acad. Publ., 2000)
5. Cowan G, in *Proc. Conf. on Advanced Statistical Techniques in Particle Physics* (Eds M R Whalley, L Lyons), IPPP/02/39, Durham 2002

6. Kaipio J, Somersalo E *Statistical and Computational Inverse Problems* (New York: Springer, 2005)
7. Barlow R, in *Proc. PHYSTAT2003, SLAC, Stanford, California, September 8–11, 2003*
8. Cowan G *Statistical Data Analysis* (Oxford: Clarendon Press, 1998)
9. Blobel V, in *PHYSTAT2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding*
10. Behnke O et al. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods* (Weinheim: Wiley-VCH, 2013)
11. Span'o F *EPJ Web Conf.* **55** (2013)
12. Schmitt S, arXiv:1611.01927; *EPJ Web Conf.* **137** 11008 (2017)
13. Hansen P C *Discrete Inverse Problems — Insight and Algorithms* (Ohio, OH: SIAM, 2010)
14. Zech G, arXiv:1607.06910
15. Egorov A Yu et al. *St. Petersburg Polytech. State Univ. J. Phys. Math.* **12** (3) (2019)
16. Adye T, arXiv:1105.1160; in *Proc. of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011*, CERN-2011-006
17. Brenner L et al., arXiv:1910.14654
18. Klepser S "Reconstruction of Extensive Air Showers and Measurement of the Cosmic Ray Energy Spectrum in the Range of 1–80 PeV at the South Pole", Dissertation Dr. rer. nat. (Berlin: Mathematisch-Naturwissenschaftlichen Fakultät I Humboldt-Univ. zu Berlin, 2008)
19. Niederhausen H "Measurement of the High Energy Astrophysical Neutrino Flux Using Electron and Tau Neutrinos Observed in Four Years of IceCube Data", PhD Thesis (New York: Stony Brook Univ., 2018)
20. Hartmann S "On the unfolding of the energy spectrum measured by the HEAT extension at the Pierre Auger Observatory", Master Deg. Dissertation (Aachen: RWTH Aachen Univ., 2015)
21. Geenen H "Reconstruction of the Primary Energy Spectrum from Fluorescence Telescope Data of the Pierre Auger Observatory", Dissertation (Wuppertal: Univ. of Wuppertal, 2007)
22. Zech G, Aslan B, in *Proc. PHYSTAT 2003, Stanford, USA, September 8–11, 2003*
23. Dembinski H P, Roth M *Nucl. Instrum. Meth. Phys. Res. A* **729** 410 (2013)
24. Andreassen A et al. *Phys. Rev. Lett.* **124** 182001 (2020); arXiv:1911.09107
25. Arratia M et al. *JINST* **17** P01024 (2022); arXiv:2109.13243
26. Rosenblatt M *Ann. Math. Statist.* **27** 832 (1956)
27. Ченцов Н Н *ДАН СССР* **147** 45 (1962)
28. Gu C, Qui C *Ann. Statist.* **21** (1) 217 (1993)
29. Loftsgaarden D O et al. *Ann. Math. Stat.* **36** 1049 (1965)
30. Айзерман М А, Браверман Э М, Розоноэр Л И *Метод потенциальных функций в теории обучения машин* (М.: Наука, 1970)
31. Kuusela M "Statistical Issues in Unfolding Methods for High Energy Physics", Master Thesis (Espoo: Aalto Univ., 2012)
32. Vischia P, arXiv:2009.02913
33. MicroBooNE Collab. "MicroBooNE low-energy excess signal prediction from unfolding MiniBooNE Monte-Carlo and data", MICROBOONE-NOTE-1043-PUB MicroBooNE docdb-15587, 2018
34. Bohm G, Zech G *Introduction to Statistics and Data Analysis for Physicists* (Hamburg: Verlag Deutsches Elektronen-Synchrotron, 2017)
35. Kuusela M, Stark P B, arXiv:1512.00905
36. Höcker A, Kartvelishvili V *Nucl. Instrum. Meth. Phys. Res. A* **372** 469 (1996)
37. Schmitt S *JINST* **7** (10) T10003 (2012)
38. D'Agostini G *Nucl. Instrum. Meth. Phys. Res. A* **362** 487 (1995)
39. Bierwagen K, in *Proc. Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, PHYSTAT 2011, CERN, Geneva, Switzerland, 17–20 January 2011* (Eds H B Prosper, L Lyons) CERN-2011-006
40. D'Agostini G, arXiv:1010.0632
41. Kuusela M, Panaretos V M *Ann. Appl. Stat.* **9** 1671 (2015)
42. Baroń P *Acta Phys. Polon. B* **51** 1241 (2020)
43. Choudalakis G, arXiv:1201.4612
44. Kartvelishvili V, in *Proc. Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011), CERN, Geneva, Switzerland, 17–20 January 2011* (Eds H B Prosper, L Lyons); CERN-2011-006
45. Richardson W H J. *Opt. Soc. Am.* **62** (1) 55 (1972)
46. Lucy L B *Astron. J.* **79** 745 (1974)
47. Zech G *Nucl. Instrum. Meth. Phys. Res. A* **716** 1 (2013)
48. Licciardi M, Quilain B, arXiv:2101.01096
49. Офф. сайт пакета ROOT, <https://root.cern.ch/>
50. Офф. сайт пакета RooUnfold, <http://hepunix.rl.ac.uk/~adye/software/unfold/RooUnfold.html>
51. Lavička R "Ultra-Peripheral Collisions at ALICE", Dissertation Thesis (Prague: Czech Technical Univ. in Prague, 2021); CERN-THESIS-2021-111
52. Тихонов А Н *ДАН СССР* **151** 501 (1963); Tikhonov A N *Sov. Math. Dokl.* **4** 1035 (1963)
53. Adye T, in *Proc. Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, PHYSTAT 2011, CERN, Geneva, Switzerland, 17–20 January 2011* (Eds H B Prosper, L Lyons) p. 313; CERN-2011-006
54. Schmelling M *Nucl. Instrum. Meth. Phys. Res. A* **340** 400 (1994)
55. Narayan R, Nityananda R *Annu. Rev. Astron. Astrophys.* **24** 127 (1986)
56. Dembinski H P, Roth M, in *Proc. Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011), CERN, Geneva, Switzerland, 17–20 January 2011* (Eds H B Prosper, L Lyons) p. 285; CERN-2011-006
57. Green P J, Silverman B W *Nonparametric Regression and Generalized Linear Models* (London: Chapman and Hall, 1994)
58. Lee T C M *Comput. Statistics Data Analysis* **42** 139 (2003)
59. Volobouev I, arXiv:1408.6500
60. Veklerov E, Llacer J *IEEE Trans. Med. Imaging.* **6** (4) 313 (1987)
61. Blobel V, OPAL Technical Note TN361 (1996)
62. Press W H et al. *Numerical Recipes in FORTRAN: The Art of Scientific Computing* 2nd ed. (Cambridge: Cambridge Univ. Press, 1992)
63. Takiya C et al. *Nucl. Instrum. Meth. Phys. Res. A* **523** 186 (2004)
64. Tang W et al. *JINST* **12** P10002 (2017)
65. Malaescu B, arXiv:0907.3791; LAL 09-107
66. Malaescu B, arXiv:1106.3107; CERN-PH-EP-2011-111
67. Malaescu B, in *Lectures for PhD students, Geneva, 2018*
68. Lyons L, in *Proc. Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011), CERN, Geneva, Switzerland, 17–20 January 2011* (Eds H B Prosper, L Lyons) p. 225; CERN-2011-006.225
69. Anykeyev V, Spiridonov A, Zhigunov V *Nucl. Instrum. Meth. Phys. Res. A* **322** (1992)
70. "Statistical Methods in Particle Physics", Heidelberg Univ., WS 2020/21, <https://www.physi.uni-heidelberg.de/~reygers/lectures/2020/smipp/>
71. Stahlman J M, Ph.D. Thesis (2014) at Publicly Accessible Penn Dissertations 1455
72. Adriani O et al. (CALET Collab.) *Phys. Rev. Lett.* **125** 251102 (2020); arXiv:2012.10319
73. van Cittert P H Z. *Phys.* **69** 298 (1931)
74. Burger H C, van Cittert P H Z. *Phys.* **79** 722 (1932)
75. Gold R, Report ANL-6984 (Lemont, IL: Argonne National Laboratory, 1964)
76. Ter-Antonyan S V *Astropart. Phys.* **28** 321 (2007); arXiv:0706.4087
77. Morhac M *Nucl. Instrum. Meth. Phys. Res. A* **559** 119 (2006)
78. Eichstädt S et al. *Metrologia* **50** 107 (2013)
79. Simon A et al. *J. High Energy Phys.* **2021** 146 (2021)
80. Li M et al. *Res. Astron. Astrophys.* **19** 145 (2019)
81. Doroshenko J J et al. *Nucl. Technol.* **33** 296 (1977)
82. Sanna R S, Technical Report EML-394, Environmental Measurements Laboratory, August 1981
83. McElroy W W et al., Technical Report AFWL-TR-67-41, US Air Force Weapons Laboratory, 1967
84. Routti J T, Sandberg J V *Radiat. Prot. Dosim.* **10** (1–4) 103 (1985)
85. Reginatto M, Goldhagen P, Technical Report EML-595, Environmental Measurements Laboratory, June 1998
86. Chen Y H et al. *Sci. China Phys. Mech. Astron.* **57** 1885 (2014)

87. Boszon A S "Measurements of Hadronic $\bar{t}\bar{t}$ Differential Cross Sections with ATLAS and Unfolding with Gaussian Processes", PhD Thesis (London: Univ. of London, 2020)
88. Koch L *JINST* **14** P09013 (2019); arXiv:1903.06568
89. Held A, Analysis Systems Typical Workshop Report, New York, June 19–20, 2019; <https://indico.cern.ch/event/822074/contributions/3471458/>
90. D'Agostini G, slides for report at Alliance Workshop on Unfolding and Data Correction, Hamburg, Germany, 27–28 May 2010; <https://wwwroma1.infn.it/dagos/unf2.hh.pdf>
91. Datta K, Kar D, Roy D, arXiv:1806.00433
92. Gagunashvili N D, arXiv:1004.2006
93. Glazov A, arXiv:1712.01814; DESY-17-214
94. Isildak B, arXiv:2001.10814
95. Goodfellow I J et al., arXiv:1406.2661
96. Butter A, Plehn T, arXiv:2008.08558
97. ARU project website, <https://aru.hepforge.org/>
98. de Boor C A *A Practical Guide to Splines* (New York: Springer-Verlag, 1978)
99. Komiske P T, Metodiev E M, Thaler J J. *High Energy Phys.* **2019** (01) 121 (2019); arXiv:1810.05165
100. Zaheer M et al., in *Advances in Neural Inf. Proc. Systems 30: Annual Conf. on Neural Information Processing Systems 2017, 4–9 December 2017, Long Beach, CA, USA* (2017)
101. Cranmer K, Pavez J, Louppe G, arXiv:1506.02169
102. Bothmann E, Del Debbio L J. *High Energy Phys.* **2019** 33 (2019)
103. Brenner L et al. *Int. J. Mod. Phys. A* **35** 2050145 (2020)
104. Milke N et al. *Nucl. Instrum. Meth. Phys. Res.* **697** 133 (2013)
105. Bourbeau J, Hampel-Arias Z J. *Open Source Software* **3** (26) 741 (2018)
106. Komiske P T, Metodiev E M, EnergyFlow package, 2019, <https://energyflow.network/>
107. Komiske P T, Metodiev E M, OmniFold package, 2021, <https://github.com/pkomiske/OmniFold>
108. Lowry K A, Johnson T L *Health Phys.* **47** 587 (1984)
109. De Sousa Lacerda M A et al., in *Proc. of the 18th Intern. Symp. on Solid State Dosimetry, Oaxaca, Mexico, 24–28 Sep. 2018*
110. Sweezy J, Hertel N, Veinot K *Nucl. Instrum. Meth. Phys. Res. A* **476** 263 (2002)
111. Pivk M, BABAR-THESIS-03/012 (2003)
112. Pivk M, Le Diberder F R *Nucl. Instrum. Meth. Phys. Res. A* **555** 356 (2005)
113. Aaij R et al. *J. High Energy Phys.* **2022** (01) 065 (2022)
114. Anderlini L, arXiv:2110.07925
115. Mathad A et al. *JINST* **16** (06) P06016 (2021)
116. Артамонов А В "Исследование рождения $\chi(nS)$ мезонов в pp -взаимодействиях при $\sqrt{s} = 7$ и 8 ТэВ в эксперименте LHCb", Дисс. ... канд. физ.-мат. наук (Протвино: Ин-т физики высоких энергий имени А.А. Логанова НИЦ "Курчатовский ин-т", 2019)
117. Andreev V et al. (H1 Collab.), arXiv:2108.12376; DESY 21-130
118. Wagner W *Mod. Phys. Lett. A* **25** 1297 (2010)
119. Gresham M I, Kim I-W, Zurek K M *Phys. Rev. D* **83** 114027 (2011)
120. Abazov V M et al. (D0 Collab.) *Phys. Rev. Lett.* **101** 191801 (2008)
121. Aaltonen T et al. (CDF Collab., D0 Collab.) *Phys. Rev. D* **97** 112007 (2018)
122. Prokhorova D S, Andronov E V J. *Phys. Conf. Ser.* **1690** 012134 (2020)
123. Evans L R, Bryant P *JINST* **3** S08001 (2008)
124. Kohn F, Ph.D. Thesis (Göttingen, 2012)
125. ATLAS Collab. *Eur. Phys. J. C* **72** 2039 (2012)
126. Chatrchyan S et al. (CMS Collab.) *Phys. Lett. B* **709** 28 (2012)
127. Savitskiy M "Measurements of differential cross sections for $\bar{t}\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons with the CMS experiment", Dissertation (Hamburg: Univ. of Hamburg, 2018)
128. Biondi S *Eur. Phys. J. Conf.* **137** 11002 (2017)
129. Sirunyan A M et al. (CMS Collab.) *J. High Energy Phys.* **2021** (03) 257 (2021)
130. Wagner-Kuhr J, arXiv:1606.02936
131. Komiske P, McCormack W P, Nachman W *Phys. Rev. D* **104** 076027 (2021); arXiv:2105.09923
132. Bellagente A et al. *SciPost Phys.* **9** 074 (2020)
133. Herrmann T "Study of Different Unfolding Methods of Kinematic Distributions of the $WZ \rightarrow WZ$ Scattering with Data and Simulation of the ATLAS Detector at the LHC", Dissertation (Dresden: Technical Univ. of Dresden, 2017)
134. Агапов Н Н и др. *УФН* **186** 405 (2016); Agapov N N et al. *Phys. Usp.* **59** 383 (2016)
135. Kolesnikov V et al. *Phys. Part. Nucl. Lett.* **16** 1, 6-15 (2019)
136. Бутенко А В и др. *УФН* **193** 206 (2023); Butenko A V et al. *Phys. Usp.* **66** 195 (2023)
137. Drnonyan J et al., report at The Conference "RFBR Grants for NICA", Evaluation of prospects for hypernuclei studies with MPD at NICA — JINR, Dubna, Russia, 20–23 October 2020
138. Gerakiev N, report at Workshop on analysis techniques for centrality determination and flow measurements at FAIR and NICA (FANI-2020) Anisotropic flow of Λ -hyperons in MPD@NICA. — NRNU "MEPhI", Moscow, 24–28 August 2020; <http://indico.oris.mephi.ru/event/181/session/1/contribution/16/>
139. Abe K et al. *Astrophys. J.* **822** 65 (2016)
140. Adriani O et al. *Phys. Rev. Lett.* **111** 081112 (2013)
141. Adriani O et al. *Phys. Rev. Lett.* **105** 121101 (2010)
142. Adriani O et al. *Astrophys. J.* **810** 142 (2015); arXiv:1512.01079
143. Adriani O et al. *Science* **332** 69 (2011)
144. Casaus J J. *Phys. Conf. Ser.* **631** 012046 (2015)
145. Ghelfi A, in *Cosmic Rays & their Interstellar Medium Environment (CRISM-2014), June 2014, Montpellier, France, Proceedings of Science, CRISM2014*, p. 013
146. Aguilar M et al. *Phys. Rev. Lett.* **126** 041104 (2021)
147. Wang Z "Measurement of Cosmic Ray Proton + Helium Flux with the DAMPE Experiment", Dissertation (Gran Sasso Science Institute, 2020)
148. Verzi V, Ivanov D, Tsunesada Y *Prog. Theor. Exp. Phys.* **2017** (12) 12A103 (2017)
149. Ivanov D *Energy Spectrum Measured by the Telescope Array Surface Detector* Doc. Diss (New Brunswick, New Jersey, 2012)
150. Curtef V *A new unfolding method for the MAGIC telescope* Doctor deg. Thes. (University of Dortmund, 2008)
151. Albert J et al. *Nucl. Instrum. Meth. A* **583** 494 (2007); arXiv:0707.2453
152. Oberndörfer M "Bayesian Unfolding of H.E.S.S. energy spectra" Master deg. Diss. (Univ. of Erlangen-Nuremberg, 2017)
153. Fei Li et al. *Results Phys.* **13** 102211 (2019)
154. Loparco F, Mazziotta M N, arXiv:0912.3695
155. Mazziotta M N, arXiv:0912.1236
156. Abdo A A et al. *Astrophys. J. Suppl. Ser.* **187** (2) (2010)
157. Mazziotta M N, arXiv:0907.0638
158. Sandroos J, arXiv:1909.07174; PoS-ICRC2019-999
159. Abbasi R *Phys. Rev. D* **83** 012001 (2011)
160. Schüssler F on behalf of the ANTARES Collab. *EPJ Web Conf.* **121** 05002 (2016)
161. Trzebinski M, Staszewski R, Chwastowski J *ISRN High Energy Physics* **2012** 491460 (2012)
162. Wang Z et al. *Nucl. Technol.* **168** 610 (2009)
163. Reginatto M, Goldhagen P, Neumann S *Nucl. Instrum. Meth. Phys. Res. A* **476** 242 (2002)
164. An F P et al. (Daya Bay Collab.) *Chinese Phys. C* **45** (7) 073001 (2021); arXiv:2102.04614
165. Zhu N M *IEEE T Nucl. Sci.* **66** 2265 (2019)
166. Peterson J H J. *Instrum.* **16** C09032 (2021)
167. Aartsen M G J. *Instrum.* **9** P03009 (2014)
168. Zinchenko A, Chabratova G *Nucl. Instrum. Meth. Phys. Res. A* **502** (2–3) 778 (2003)
169. Blobel V, Kleinwort C, hep-ex/0208021; DESY 02-077
170. Reginatto M *Radiat. Meas.* **45** 1323 (2010)
171. Dommert M *Current Directions Biomed. Eng.* **3** (2) 83 (2017)
172. Zimbal A, at ECPD 2015 – 1st EPS Conf. on Plasma Diagnostics – Frascati – Proc. of Science (ECPD2015)
173. Venanzoni G *AIP Conf. Proc.* **1182** 665 (2009)
174. Ablikim M *Chinese Phys. C* **37** 063001 (2013)
175. Samanta S *Nucl. Phys. A* **1005** 121896 (2021); arXiv 2002.12235
176. Aguilar M et al. (AMS Collab.) *Phys. Rev. Lett.* **114** 171103 (2015)
177. Green D M, PhD Thesis (Univ. of Maryland, 2016)

178. Adriani O et al. *Astrophys. J.* **791** 93 (2014)
179. An Q et al. *Sci. Adv.* **5** (9) eaax3793 (2019)
180. Adriani O et al. *Phys. Rev. Lett.* **126** 241101 (2021)
181. Aab A et al. *Phys. Rev. D* **102** 062005 (2020)
182. Zanin R "Observation of the Crab pulsar wind nebula and micro quasar candidates with MAGICz", Dissertation (Autonomous Univ. of Barcelona, 2011)
183. Albert J et al. *Astrophys. J.* **663** 125 (2007)
184. Aartsen M G et al. (IceCube Collab.) *Eur. Phys. J. C* **77** (10) 692 (2017)
185. Aguilar J A et al. (IceCube Collab.) *Eur. Phys. J. Part. Fields* **75** (3) 116 (2015)
186. Richard E et al. *Phys. Rev. D* **94** 052001 (2016)
187. "Statistical methods 2021", Institute of Particle and Nuclear Physics, <https://ipnp.cz/?page.id=4280>
188. Логащенко И Б *Методы анализа экспериментальных данных. Электронный лекционный курс* (Новосибирск: НГУ, 2013)
189. Логащенко И Б, Эйдельман С И *УФН* **188** 540 (2018); Logashenko I B, Eidel'man S I *Phys. Usp.* **61** 480 (2018)
190. Alekseev V V et al. *J. Phys. Conf. Ser.* **1390** 012071 (2019)
191. Baron P, arXiv:2001.05877
192. Wei X et al. *IEEE Trans. Signal Process.* **70** 2962 (2022); arXiv:2107.02848
193. Pop F *Adv. High Energy Phys.* **2014** 507690 (2014)
194. Kuusela M, in *PhyStat-v 2019*, CERN, Geneva, Switzerland
195. Kuusela M, CP3 Remote Seminar report (Université Catholique de Louvain), July 1, 2020, <https://agenda.irmp.ucl.ac.be/event/4000/>

Review of unfolding methods

Yu.V. Bogomolov^(1,2,a), V.V. Alekseev^(1,2), O.A. Levanova^(1,2,b), A.G. Mayorov⁽²⁾, V.V. Malakhov⁽²⁾

⁽¹⁾ Demidov Yaroslavl State University, ul. Sovetskaya 14, 150003 Yaroslavl, Russian Federation

⁽²⁾ National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),

Kashirskoe shosse 31, 115409 Moscow, Russian Federation

E-mail: ^(a) mathematics@inbox.ru, ^(b) olaydy@gmail.com

Very high accuracy and sensitivity have become attainable by modern instruments for experimental measurements of physical quantities in various scientific fields. Yet it is still impossible to completely eliminate the influence of instrumental effects on the result. The measured values of a physical quantity inevitably differ, sometimes significantly, from the true ones. The question therefore arises of restoring the true distributions from the measured ones, taking the specific features of the experiment and the characteristics of scientific instruments into account. Different approaches are in use based on a mathematical model of the instrument and the formulation of the deconvolution problem. We describe this problem, key ideas and methods for its solution, and features and implementation details using the example of elementary particle physics and space physics experiments.

Keywords: unfolding, deconvolution, spectrum restoration methods, statistical estimates, Bayesian methods, regularization, binning-free methods, machine learning

PACS numbers: **02.50. – r**, **02.60. – x**, 96.50.S–, 96.50.sb

Bibliography — 195 references

Received 23 January 2022, revised 7 April 2022

Uspekhi Fizicheskikh Nauk **193** (6) 669 – 685 (2023)

Physics – Uspekhi **66** (6) (2023)

DOI: <https://doi.org/10.3367/UFNr.2022.05.039189>

DOI: <https://doi.org/10.3367/UFNe.2022.05.039189>